EPIDEMIOLOGY AND GENOMIC CHARACTERIZATION OF PORCINE *STREPTOCOCCUS SUIS*

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
UNIVERSITY OF MINNESOTA
BY

April Aquino Estrada

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Advisors: Dr. Connie J. Gebhart & Dr. Douglas G. Marthaler

August 2021

ACKNOWLEDGEMENTS

I would like to thank all my family and friends who have supported me in all aspects of life. To my family and friends in California, thank you for sending motivational texts, calling to ask how I'm doing, and for spending time with me during my unexpected trips to CA. To my brother in Minnesota, thank you for visiting often and being my emotional support, especially during the first few years when I was living alone. With enormous love and gratitude, I want to thank my husband Alex. He has motivated me since high school by reminding me of my potential and ability to succeed. With his unwavering support, I was able to take risks, such as attending grad school, and follow my passions.

I am grateful to my advisors Dr. Connie Gebhart and Dr. Douglas Marthaler. As a first-generation college student, I had many questions and concerns about grad school. Connie and Doug gave me the opportunity to do research and were extremely patient and supportive of my learning and growth. Connie, thank you for your guidance over the past years and for challenging me to become an independent researcher. Doug, thank you for the continuous engagement and support, especially in learning bioinformatics and becoming a better academic writer. Thank you to the other committee members, Dr. Mark Rutherford, Dr. Tim Johnson, Dr. Albert Rovira, and Dr. Stephanie Rossow for their insightful comments and challenging questions.

Thanks to Dr. Fabio Vannucci and my lab mates, Lacey, Erika, Talita, and Marta for being so welcoming and enthusiastic. Also, thanks to Lacey, Marta, and the VDL bacteriology lab for all their technical assistance over the past 5 years.

I'd also like to recognize the faculty, staff, and graduate students that have contributed to both my academic and personal growth. Annie, thank you for being a great friend and a fellow nerd in class and at seminars and conferences. Finally, I'd like to thank all my funding sources for making this research possible and for exposing me to professional development opportunities.

DEDICATION

Para mí mamá y papá, Marina y Avelino.

Por siempre ser orgullosos de mí y por darme el apoyo, el animo, y la confianza que necesito para cumplir mis sueños. Gracias por todos los sacrificios que hacen por mí. Los quiero mucho.

ABSTRACT

Streptococcus suis is a significant cause of systemic and central nervous system (CNS) disease in pigs and is a welfare and economic issue worldwide. The control of *S. suis* disease in the United States is hindered by the lack of contemporary genotypic and epidemiological data. Lacking is information on the genomic differences between pathogenic or disease-causing strains and commensal strains. The overall goal of this dissertation was to investigate the distribution and population structure of *S. suis* and determine the genomic characteristics and virulence-associated factors of pathogenic strains to better understand the observed persistence and severity of *S. suis* infections in the United States. Serotyping and multilocus sequence type (MLST) are traditional methods for characterizing *S. suis*. U.S. isolates represented uniquely prevalent subtypes with several subtypes being associated with disease as determined by pathotype classifications. These results increased the knowledge about the serotype distribution and population structure of *S. suis* in the United States and established a link between subtype and pathotype to better predict pathogenicity. Genomic prediction of previously published virulence-associated genes (VAGs) of *S. suis* illustrated a high number of VAGs found across pathotypes indicating pathogenic and commensal strains share many VAGs. Statistical analysis of VAGs and pathotype classifications confirmed classical virulence markers of Eurasian strains are not sufficient for identifying pathogenic strains in the U.S. Regression analysis was used to predict the association of published VAGs of *S. suis* and pathotype. Results indicated that VAG profiling should be complemented with the two newly proposed VAGs. This supports previous studies that North American and Eurasian pathogenic strains are genetically different. The same MLST-VA genotype patterns were identified in multiple production companies, suggesting these patterns may be widespread as opposed to originating from a common source. Comparative pan-genome analysis of *S. suis* isolates illustrated genomic regions (accessory genome) associated with virulence-related functions. Further analysis of the accessory genome by pathotype revealed three novel candidate VAGs of *S. suis* with potential virulence-related functions that may better serve as predictive markers of pathogenicity. Based on *in silico*

analysis, high resistance to tetracyclines and erythromycin were observed regardless of pathotype, and mobile genetic elements appeared to play a limited role in the transfer of antimicrobial resistance genes in *S. suis*. In summary, this dissertation investigated the genomic diversity of *S. suis*, the subtyping of pathogenic and commensal strains, and the identification and potential epidemiological surveillance of clones within and between swine production companies. The information provided by our work can be applied to diagnostics, epidemiological monitoring, and control of *S. suis*.

Supplementary information:

Additional file 4.1. Gene clusters (n=8,231) identified in Roary analysis of *S. suis* genomes (n=208). Binary matrix representing the presence (1) and absence (0) of all 8,231 gene clusters identified by the Roary analysis performed on all 208 *S. suis* genomes. **[.xlsx]**

Additional file 4.2. COG functional classifications for the 995 pan-gene clusters that met the criteria. Results of the BLAST searches against the Predicted Cluster of Orthologous Groups (COG) database. Classification as core or accessory and present (yes) or absent (no) in each pathotype is listed. **[.xlsx]**

Additional file 4.3. Gene clusters (n=6,407) identified in Roary analysis of *S. suis* genomes (n=161). Binary matrix representing the presence (1) and absence (0) of all 6,407 gene clusters identified by the Roary analysis performed on the 161 *S. suis* genomes representing only the pathogenic and commensal pathotypes. **[.xlsx]**

Additional file 4.4. ARGs, VAGs, and MGEs identified in *S. suis* genomes (n=208). Binary matrix representing the presence (1/yes) and absence (0/no) of ARGs, MGEs, MGE-associated ARGs, and MGE-associated VAGs. **[.xlsx]**

Additional file 4.5. Prokka annotations of MGE drafts. List of prokka annotations of the plasmid, ICE, IME, composite transposon, and 89K PAI draft sequences identified in this study. **[.xlsx]**

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ARG, antimicrobial resistance gene

CC, clonal complex

CI, confidence interval

COG, Clusters of Orthologous Groups of proteins

EF/*epf*, extracellular protein factor

ICE, integrative conjugative element

IME, integrative mobilizable element

Inf, infinite

MGE, mobile genetic element

MLST/ST, multilocus sequence typing/sequence type

LASSO, least absolute shrinkage and selection operator

MRP/*mrp*: muramidase-released protein

MCA, multiple correspondence analysis

NT, nontypeable

OR, odds ratio

ofs, serum opacity factor

PAI, pathogenicity island

SLY/*sly*, suilysin

SNP, single nucleotide polymorphism

srtF, sortase F

SRST2, short read sequence typing for bacterial pathogens program

VAG or VA genotype, virulence-associated gene, virulence-associated genotype

WGS, whole-genome sequencing

CHAPTER 1
Introduction

1.1 General Introduction

Streptococcus suis is a bacterial pathogen which primarily occurs in nursery and recently weaned pigs. Systemic disease is characterized by meningitis, arthritis, endocarditis, pneumonia, and septicemia causing great economic losses and welfare concerns to the swine industry worldwide (Gottschalk and Segura, 2019). *S. suis* is also an important human pathogen and disease is characterized by meningitis, septicemia, and outbreaks of streptococcal toxic shock syndrome (Lun et al., 2007). Most pigs are colonized early in age by *S. suis* and serve as carriers of *S. suis* which can then be transmitted between animals or to humans (Torremorell et al., 1998). However, only certain strains are pathogenic or capable of causing disease, thus the identification and characterization of *S. suis* contributes to our understanding of strains associated with disease.

Characterization of *S. suis* by serotyping and multilocus sequence typing (MLST) revealed that serotype 2 (SS2) and sequence type (ST1) are the predominant subtypes involved in swine clinical cases, globally, but differences in distributions and prevalence of particular subtypes depend on the geographical location (Goyette-Desjardins et al., 2014). A majority of the reported data on serotype and ST distributions in North America is from Canada with only a handful of studies reporting on isolates from the United States (Goyette-Desjardins et al., 2014). Therefore, an update on the characterization and genetic diversity of U.S. isolates is needed. Chapter 2 of this dissertation focuses on the serotype and MLST characterization of 208 *S. suis* isolates from the United States and utilization of a pathotype classification system to determine prevalent pathogenic subtypes.

Many virulence-associated factors (factors and markers) have been proposed for *S. suis* and were predominantly identified in studies using SS2 strains (Fittipaldi et al., 2012; Segura et al., 2017). Well-studied virulence-associated factors include the capsular polysaccharide (CPS) and extracellular protein factor (EF), muramidase-released protein (MRP), and suilysin (SLY) which are markers for highly virulent Eurasian SS2 strains. North American SS2 strains are genetically different (lack one or more of these markers) and less virulent than Eurasian strains (Fittipaldi et al., 2011; Auger et al., 2016a). However, the distribution of these markers and other virulence-associated factors in North America, particularly the United States, and the correlation to pathogenicity remain poorly known (Fittipaldi et al., 2009, 2011). Chapter 3 focuses on the genomic characterization of *S. suis* in order to determine published virulence-associated genes associated with systemic and central nervous system (CNS) disease. The genetic relatedness of U.S. isolates was also investigated to identify associations between virulence-associated genotypes and subtypes, for epidemiological purposes, and for guiding vaccine strain selection.

S. suis virulence-associated factors that are important for pathogenicity are poorly understood, especially for non-SS2 strains, may not be consistently involved in virulence, and/or are geographically specific (Fittipaldi et al., 2012; Segura et al., 2017). Thus, novel candidate virulence-associated factors should be investigated using a diverse set of isolates (e.g., multiple serotypes, STs). Chapter 4 of this dissertation focuses on the comparative genomic analysis of *S. suis* to determine novel candidate virulence-associated genes, antimicrobial resistance genotypes, and other genomic characteristics of pathogenic isolates of different serotypes and STs.

1.2 Literature Review

1.2.1 *Streptococcus* genus

Streptococcus taxonomy

Streptococcus is a genus of gram-positive, encapsulated, spherical (cocci) bacteria. Currently, the *Streptococcus* genus is composed of over 65 species classified

into six to eight main “species groups”, the anginosus, bovis, downei, mitis, mutans, pyogenic, salivarius, and sanguinis groups, and several ungrouped *Streptococcus* species, including *S. suis* (Facklam, 2002; Richards et al., 2014). Initial classification and differentiation of *Streptococcus* species were based on morphological, cultural, biochemical, and serological properties (Facklam, 2002; Gobbetti and Calasso, 2014). Early phylogenetic studies on prokaryotic taxonomy utilized DNA-DNA hybridization and 16S rRNA sequencing to establish and differentiate bacterial species, and strains demonstrating at least 70% DNA-DNA relatedness and 97-99% 16S rRNA identity are generally considered members of the same species (Stackebrandt and Goebel, 1994). The taxonomy of the *Streptococcus* genus continues to be revised as molecular techniques evolve.

There are a total of 35 known serotypes (SS) of *S. suis* (1-34 and 1/2). 16S rRNA analysis of *S. suis* serotypes 1-34 and 1/2 demonstrated a divergence between serotypes 20, 22, 26, 32, 33, and 34 and the other 29 serotypes (Chatellier et al., 1998). Inclusion of other *Streptococcus* species further illustrated that SS32, SS33, and SS34 were more related to other members of the genus than to the other *S. suis* serotypes (Chatellier et al., 1998). However, the biochemical profiles of all 35 serotypes were similar. SS32 and SS34 were later reclassified as *Streptococcus orisratti* based on 16s rRNA and chaperonin-60 (cpn60) genetic analyses (Hill et al., 2005). SS20, SS22, and SS26 were reclassified as *Streptococcus parasuis* based on DNA-DNA relatedness, *sodA* and *recN* phylogenies, and whole-genome sequencing (Tien et al., 2013; Nomoto et al., 2015). SS33 was also removed from the *S. suis* taxon based on molecular and biochemical analyses and proposed as *Streptococcus ruminantium* (Tohya et al., 2017).

Evolution of the Streptococcus genus

Whole-genome sequencing of streptococcal species has revealed important aspects of the evolution of this genus. *Streptococcus* genomes are relatively small (~2 Mb) compared to other bacteria, have a low GC content (average 39%), and are comprised of an average of 2,000 protein coding genes (Lefébure and Stanhope, 2007; Gao et al., 2014). Further, the different *Streptococcus* species share a small percentage of

predicted genes, which varied depending on the number of genomes compared. Analysis of 138 *Streptococcus* genomes representing 42 species and all species groups indicated only about 9% of the predicted gene content is shared across species and comprises the *Streptococcus* core genome (Gao et al., 2014). This high species- and strain-specific diversity is not surprising because the genus *Streptococcus* is composed of both pathogenic and commensal bacteria found in a wide range of hosts and niche environments (Ruoff et al., 2003; Anisimova et al., 2007).

Evolutionary mechanisms in bacteria include positive (adaptive) selection, a process in which advantageous mutations increase in frequency within a population over time. Genomic analysis of 26 *Streptococcus* genomes representing six different species (*S. agalactiae*, *S. mutans*, *S. pneumoniae*, *S. pyogenes*, *S. suis*, and *S. thermophilus*) demonstrated that up to 40% of core genes are under positive selection and selection pressure was greater between species than within a species (Lefébure and Stanhope, 2007). This pattern suggests positive selection plays a larger role in species differentiation than strain differentiation. Functional analyses of genes under positive selection in bacterial pathogens revealed an enrichment of genes encoding cell surface/membrane and metabolic proteins (Chen et al., 2006; Petersen et al., 2007; Cao et al., 2017). These proteins included putative virulence factors associated with bacterial evasion of host immune responses, supporting the role of selection pressure in the evolution of pathogen-host interactions. *S. suis* appears to be under strong selection pressure which may be related to host-adaptation and its ability to infect both pigs and humans (Lefébure and Stanhope, 2007).

Recombination is a major driver of strain evolution in streptococcal pathogens, such as *S. pneumoniae*, and the rate at which recombination occurs varies within the genus (Lefébure and Stanhope, 2007; Andam and Hanage, 2015). Horizontal (or lateral) gene transfer introduces novel genes, typically accessory genes not conserved across all members of a species, into a host genome via recombination (homologous or site-specific) (Lawrence and Retchless, 2009; Blakely, 2015). The lateral movement of genes can occur within a species or between species, both closely and distantly related, enabling

the evolution of genetic traits that allow bacteria to adapt to new niche environments (Blakely, 2015). Furthermore, newly acquired genes may confer traits necessary for a pathogenic lifestyle, such as virulence factors and antimicrobial resistance determinants acquired through mobile genetic elements (Jackson et al., 2011).

1.2.2 *Streptococcus suis* disease

Clinical signs & epidemiology

S. suis infections mostly occur in nursery and recently weaned pigs between 5 and 10 weeks of age and clinical signs include fever, lameness, swollen joints, neurological signs, and sudden death (Reams et al., 1994; Gottschalk and Segura, 2019). Clinical signs of central nervous system (CNS) disease or meningitis are head tilt, ataxia, paddling, and convulsing. Clinical signs of respiratory disease, such as coughing and "thumping", have been observed in *S. suis* infected pigs (Reams et al., 1994), however, it is unclear whether *S. suis* is a primary cause of respiratory disease because *S. suis* is part of the porcine respiratory disease complex along with other respiratory pathogens (Obradovic et al., 2021). Pigs can become infected through vertical transmission of *S. suis* from infected sow to piglets during farrowing (Amass et al., 1997) or horizontal transmission due to direct contact (nose-to-nose) with infected or carrier pigs (Dekker et al., 2013) with the latter being the predominant route. Risk factors associated with *S. suis* disease include temperature variations (Zou et al., 2018), high pig density (Hughes et al., 2009), and poor air quality (Bonifait et al., 2014). The potential aerosolization of *S. suis* and airborne transmission in pigs has been demonstrated for SS2 and SS1/2 (Berthelot-Hérault et al., 2001; Bonifait et al., 2014).

Zoonosis

S. suis is a zoonotic pathogen and causes meningitis and septicemia in humans (Lun et al., 2007). Two outbreaks of human infection in China were characterized by a streptococcal toxic shock-like syndrome with high mortality (Tang et al., 2006) which had only been previously associated with *S. pyogenes* (group A streptococcus) infection. The first human infection was reported in 1968, and over 1600 reported human cases

have occurred, mainly in Asian countries like China, Vietnam, and Thailand (Goyette-Desjardins et al., 2014; Huong et al., 2014). Infections in Asia are mainly associated with the consumption of raw pork products but also occur through occupational or household exposure to pigs or contaminated pork (Hughes et al., 2009; Huong et al., 2014). Human infections in Western countries are primarily associated with occupational hazard (abattoir workers, butchers, and pig breeders) and are rare due to differences in food consumption practices between Western and Asian countries (Arends and Zanen, 1988; Gottschalk et al., 2010).

Pathogenesis

Multiple routes of natural transmission of *S. suis* have been proposed with the oro-nasal route widely accepted as the primary route of infection in pigs (Dekker et al., 2013). Commensal strains naturally colonize the upper respiratory tract (tonsils and nasal cavities), genital, and intestinal tract of pigs without causing clinical disease (Arends et al., 1984; Gottschalk and Segura, 2019). However, *S. suis* is an opportunistic pathogen and can cause disease under favorable conditions, such as mixed infections with viral (Porcine Reproductive and Respiratory Syndrome Virus, Swine Influenza Virus) or other bacterial pathogens (*Pasteurella multocida*, *Mycoplasma hyopneumoniae*) (Vötsch et al., 2018; Obradovic et al., 2021). Virulent strains cause disease by colonizing the respiratory tract, penetrating host mucosal barriers, spreading through the bloodstream, and invading different organs.

Physical barriers against *S. suis* colonization include a mucus layer covering the upper respiratory tract and the epithelial cell layer which capture and clear bacterial cells. Previous studies on the human pathogen *Streptococcus pneumoniae* suggest that the capsule reduces entrapment of cells in this mucus layer and its immune cell components, enhancing colonization and survival in the respiratory tract (Nelson et al., 2007). *S. suis* is a well-encapsulated bacterium and there is evidence that a similar mechanism for limiting mucus-mediated clearance is present in *S. suis* (Seitz et al., 2014). Adhesion of bacterial cells to host cells typically requires receptor-ligand interactions or the binding of bacterial surface components (adhesions) by host receptors (Di Martino, 2018). The *S. pneumoniae*

capsule partially hinders adherence of pneumococci to epithelial cells by reducing exposure of bacterial adhesin molecules embedded within the cell wall (Hammerschmidt et al., 2005; Novick et al., 2017). A similar pattern of reduced adherence by encapsulated strains was observed in *S. suis* (Lalonde et al., 2000; Esgleas et al., 2005). Furthermore, there appear to be changes in capsule thickness during different stages of infection with *S. suis* possessing a thinner capsule during colonization (Gottschalk and Segura, 2000). These findings suggest dynamic regulation of capsule expression during the infectious process.

Characterized SS2 *S. suis* virulence-associated factors that contribute to adherence of epithelial cells include factor-H binding proteins (Fhb and Fhbp) (Roy et al., 2016), glyceraldehyde-3-phosphate dehydrogenase (GAPDH) (Brassard et al., 2004; Wang and Lu, 2007), and suilysin (Seitz et al., 2013; Meng et al., 2016). Previous studies indicate a low capacity for invading host epithelial cells by encapsulated *S. suis* strains (Lalonde et al., 2000; Benga et al., 2004) but the mechanisms of invasion are not well-understood (Segura et al., 2016). A proposed mechanism is suilysin-mediated invasion which suggests the suilysin toxin binds to the host epithelial cell membrane, activates the actin cytoskeleton, and promotes membrane ruffling and bacterial uptake (Seitz et al., 2013). Alternatively, damage of cells by suilysin-induced cytotoxicity contributed to invasion of porcine epithelial cells (Meng et al., 2016) and human brain microvascular endothelial cells (BMECs) (Charland et al., 2000).

The innate immune system plays a role in preventing the colonization and dissemination of bacterial pathogens. The mucosal epithelial barrier is also composed of antimicrobial molecules, such as lysozymes, cytokines, and chemokines secreted by epithelial cells and/or innate immune cells (McGuckin et al., 2015). Lysozyme resistance may contribute to successful mucosal colonization and is multifactorial and highly variable among SS1, SS2, SS7, and SS9 with SS7 and SS9 demonstrating the highest lysozyme resistance (Wichgers Schreur et al., 2012). *S. suis* has developed mechanisms for escaping inflammatory immune cells (e.g. neutrophils and dendritic cells), such as expressing the virulence-associated factors subtilisin-like protease (SspA) (Bonifait and

Grenier, 2011) and serine protease (Vanier et al., 2009) that degrade or cleave chemokines reducing the recruitment and/or activation of the immune response. Additionally, *S. suis* SS2 may promote an exacerbated release of inflammatory mediators by monocytes and macrophages, contributing to the development of sepsis and meningitis (Segura et al., 1999; Dominguez-Punaro et al., 2007). The role of the host complement system in protecting against invasive *S. suis* infection is debatable as the efficiency of complement in clearing *S. suis* varies depending on the animal model (Segura et al., 2016). Virulence-associated factors of *S. suis* can interfere with complement by inhibiting deposition of complement proteins (opsonization), inactivating the alternative complement pathways, and/or degrading complement molecules, which prevents subsequent complement-dependent killing (e.g. phagocytosis) (Fittipaldi et al., 2012; Segura et al., 2016). These factors include Fhb, CPS, suilysin, DltA, and PdgA (peptidoglycan N-deacetylation). Suilysin also protects *S. suis* against the immune response via its cytotoxic effects against monocytes and neutrophils (Segura and Gottschalk, 2002; Chabot-Roy et al., 2006). Various other virulence factors, such as proteins associated with nutrient uptake (AdcR, Zur, FeoB, TroA), have been identified *in vitro* and/or *in vivo* as contributing to survival in blood and dissemination to different organs (e.g. liver, lung, heart) (Fittipaldi et al., 2012; Arenas et al., 2019).

In the natural route of infection, *S. suis* first becomes systemic and can then develop into CNS disease. Although the mechanism for the development of meningitis in pigs in not fully understood, the hypothesis states meningeal pathogens must transverse the blood-brain barrier (BBB) and/or the blood-cerebrospinal fluid (CSF) barrier to cause meningitis (Kim, 2008; Gerber and Nau, 2010). *S. suis* SS2 strains have been shown to adhere to but not invade human umbilical vein endothelial cells (Charland et al., 2000), while being shown to adhere to, invade, and survive within porcine BMECs (Vanier et al., 2004). Multiple virulence-associated factors are involved in the interactions between *S. suis* and BBB cell types of human and porcine origin (Gottschalk and Segura, 2000; Fittipaldi et al., 2012). Fhb binds to Gb3 expressed on the surface of BMECs (Kong et al., 2017) while enolase binds to the extracellular matrix proteins fibronectin and plasminogen (Esgleas et al., 2008) both promoting adhesion to BMECs (human and

porcine, respectively) and paracellular translocation (BBB permeability) of *S. suis*. Enolase can significantly increase BBB permeability by inducing leukocyte migration via IL-8 secretion and/or porcine BMEC apoptosis (Sun et al., 2016; Liu et al., 2021). This increased recruitment of leukocytes into the CNS plays a role in the development of meningitis. The "Trojan horse" or "modified Trojan horse" theory has been proposed as a mechanism for *S. suis* to cross the BBB (Williams and Blakemore, 1990; Gottschalk and Segura, 2000). This hypothesis suggests that *S. suis* travels to the BBB while inside or surface-associated with monocytes which transverse into the subarachnoid space.

The blood-CSF barrier, formed by choroid plexus epithelial cells (CPEC) and tight junctions, is another potential site for bacterial entry into the CNS (Engelhardt and Sorokin, 2009). Multiple studies have reported lesions at the choroid plexus of infected animals (pigs and mice) (Madsen et al., 2002; Dominguez-Punaro et al., 2007), however, investigation of invasion and translocation across this barrier were limited due to the lack of appropriate *in vitro* models. Tenenbaum et al. (2009) utilized a functional *in vitro* porcine CPEC model, as determined by intact tight junction function and transepithelial electrical resistance values, and demonstrated that *S. suis* invade CPECs in a polar fashion from the basolateral side (Tenenbaum et al., 2009). Similar observations were made using a human model of the blood-CSF barrier (Schwerk et al., 2012). As with migration across the BBB, *S. suis* can disrupt barrier function through apoptosis or necrosis of porcine CPECs (Tenenbaum et al., 2005, 2006).

The gastrointestinal tract has been proposed as a secondary site of infection in piglets and a common route of infection in humans (Gottschalk and Segura, 2019). Quantification of *S. suis* within the gastrointestinal microbiota suggests *S. suis* colonizes the intestine after weaning (21 days of age) (Su et al., 2008). SS2 human and SS9 porcine clinical strains illustrated significantly higher adhesion to human intestinal epithelial cells (IEC) and porcine IECs, respectively (Ferrando et al., 2015). In addition, translocation across human EICs was associated with CC1. These results agree with previous studies reporting the zoonotic potential of SS2 but not SS9 (Goyette-Desjardins et al., 2014). SS2 is capable of translocating across the porcine gastrointestinal tract *in vivo* after oral

(Ferrando et al., 2015) or intestinal (Swildens et al., 2004) challenge, infecting multiple organs, and causing disease. The low survival rate of SS2 and SS9 strains in the porcine stomach may indicate the porcine stomach is an effective barrier against oro-gastrointestinal infections but further studies are needed (Warneboldt et al., 2016).

Treatment and prevention

Beta-lactam antibiotics such as penicillin, ampicillin, ceftriaxone, and ceftiofur, and fluoroquinolone antibiotics such as enrofloxacin are used to treat *S. suis* infections (Seitz et al., 2016). Globally, resistance to penicillin and other beta-lactam antibiotics is generally low among *S. suis* isolates, while resistance to tetracyclines has been reported in both clinically healthy and diseased pigs in the Americas, Asia, and European countries (Seitz et al., 2016; Hernandez-Garcia et al., 2017; Arndt et al., 2019; Hayer et al., 2020; Tan et al., 2021). Regional and geographical differences in antimicrobial resistance patterns of *S. suis* have been observed, suggesting the importance of monitoring antimicrobial resistance patterns in different regions (Seitz et al., 2016; Yongkiettrakul et al., 2019; Hadjirin et al., 2020). The movement towards prudent and responsible use of antimicrobials in pigs and humans may improve both animal and human health by preventing or reducing antimicrobial-resistant bacteria.

An effective cross-protective vaccine for *S. suis* is strongly needed but difficult to attain partially due to the wide genetic and phenotypic variability of *S. suis*, and vaccine efficacy is largely limited to protection against homologous challenge. Homologous protection against *S. suis* challenge has been investigated for serotypes 1, 2, 1/2, 3, 9, and 14, and few vaccine strains have been investigated in field trials (Segura, 2015). Whole-cell bacterin vaccines provide varying levels of protection against homologous challenge with most providing low or no protection against morbidity and/or mortality (Segura, 2015). One approach for eliciting a stronger immune response against bacterial cells is to expose cell wall proteins of the vaccine strain. Previous studies demonstrated a survival rate of 100% and a reduction in clinical signs in pigs vaccinated with nonencapsulated mutant strains (whether live or killed) (Wisselink et al., 2002; Fittipaldi et al., 2007), which was comparable to full protection induced by a formalin-killed virulent capsulated

strain (Wisselink et al., 2002). There is also evidence that live attenuated or avirulent strains (SS2) can confer good protection in mice (Quessy et al., 1994; Li et al., 2019) and pigs (Holt et al., 1988; Busque et al., 1997; Fittipaldi et al., 2007) against homologous challenge, although multiple immunizations were given. Only two studies, to our knowledge, have reported on the efficacy of SS9 bacterin vaccines. Dekker et al. (2012) utilized a pig model optimized for evaluating transmission and colonization (not mortality or clinical signs), and demonstrated that vaccination against a homologous SS9 strain did not reduce transmission nor colonization (Dekker et al., 2012). Buttner et al. (2012) reported that a SS9 bacterin vaccine elicited significant protection against mortality in pigs against heterologous SS9 challenge (Büttner et al., 2012). Adjuvant formulation can influence the immunogenicity and protective efficacy of a *S. suis* vaccine. Formulation of vaccines with water-in-oil emulsion adjuvants reduced both mortality and clinical signs in pigs (Holt et al., 1990; Wisselink et al., 2001; Pallarés et al., 2004) while an aluminum hydroxide-based adjuvant failed to induce good protection (Holt et al., 1990; Wisselink et al., 2001). Maternal immunity in piglets from sows immunized with autogenous vaccines is inconsistent due to highly variable vaccine strains (different serotypes), vaccination protocols, and types of challenge (homologous vs heterologous) (Segura, 2015; Corsaut et al., 2020, 2021).

Immune response to heterologous challenge is variable between and within serotypes. Immunization of pigs with a formalin-killed bacterin protected against heterologous SS2 challenge, although cross-protection could be attributed to a common MRP+/EF+ phenotype (Wisselink et al., 2001). A SS2 (MRP+EF+SLY+) bacterin vaccine failed to induce significant opsonizing antibodies or to protect against a virulent SS9 (MRP*+/SLY+) strain (Baums et al., 2009). Jiang et al. (2019), however, described the cross-protective potential (66% protection) of a SS2 subunit vaccine against SS9 infection (Jiang et al., 2019). Vaccines containing SS1/2 protected well against challenge with SS1 and SS2, but immunization with SS1 did not protect against SS2 and vice versa (Kebede et al., 1990; Foster et al., 1994). Inconsistent results have been published on the efficacy of SS2 vaccines against heterologous challenge with different STs of SS2

(Quessy et al., 1994; Aranda et al., 2009), providing further proof that protection may be strain specific and a diverse set of serotypes and STs needs to be evaluated.

Among the many putative and proposed virulence-associated factors of *S. suis*, only a few, mostly surface-associated proteins, have been evaluated for their immunogenic properties and capacity to induce cross-protection (Segura, 2015). The CPS is poorly immunogenic (Calzas et al., 2017; Goyette-Desjardins et al., 2020) but administering SS2 CPS with specific adjuvants and/or preparing a glycoconjugate improved levels of opsonizing SS2-specific anti-CPS antibodies (Elliott et al., 1980; Goyette-Desjardins et al., 2016) and provided significant protection of pigs against systemic infection (Goyette-Desjardins et al., 2016). Furthermore, characterization of monoclonal antibodies indicated that the glycoconjugate vaccine elicited a serotype-specific protective response (Goyette-Desjardins et al., 2019). Antigens conserved across many different *S. suis* serotypes may serve as cross-protective candidates for a "universal" vaccine. Conserved virulence factors of *S. suis* demonstrating good immunogenicity include MRP (Baums et al., 2009; Li et al., 2017), Sao (Li et al., 2006), enolase (Feng et al., 2009; Zhang et al., 2009; Dumesnil et al., 2019), and suilysin (Jacobs et al., 1994; Liu et al., 2009, p. 201). An MRP and EF subunit vaccine induced good protection against homologous and heterologous *S. suis* challenges in a pig model and significantly greater than MRP and EF alone, but protection was adjuvant-specific (Wisselink et al., 2001). Similar adjuvant-specific protection was observed for Sao (SS2 infection) (Li et al., 2006, 2007) and enolase (SS2 or 7 infection) (Feng et al., 2009, p. 200; Zhang et al., 2009, p. 200), but contradictory results on the effects of adjuvant formulation have also been reported (Esgleas et al., 2009; Dumesnil et al., 2019). Despite the promising results in experimental trials, development of cross-protective subunit vaccines is still a challenge because virulent strains may lack several virulence-associated factors (Gottschalk and Segura, 2000; Segura et al., 2017).

1.2.3 Molecular techniques for the characterization and subtyping of *S. suis*

Serotyping

There are a total of 35 known serotypes of *S. suis* (1-34, and 1/2), however six have been reclassified as *Streptococcus orisratti* (SS32 & SS34), *Streptococcus parasuis* (SS20, SS22, and SS26) or *Streptococcus ruminantium* (SS33) (Hill et al., 2005; Tien et al., 2013; Tohya et al., 2017). Traditionally, serotype is determined by agglutination of the *S. suis* capsule to reference antisera, thus failing to serotype non-encapsulated strains (non-typeable). Multiplex PCR assays were developed to detect serotype-specific capsular polysaccharide synthesis (*cps*) genes of *S. suis* and determine serotype using two to four sets of multiple PCR amplifications (Liu et al., 2013; Kerdsin et al., 2014; Okura et al., 2014). These multiplex PCR assays were able to correctly determine the serotypes of human and porcine field isolates and of isolates non-typeable by serological assays but are not able to distinguish between SS2 and SS1/2 and SS1 and SS14. Sequence analysis of non-typable porcine strains led to the discovery of 27 novel *cps* loci (NCLs) and the development of molecular assays for identifying these NCLs (Pan et al., 2015; Zheng et al., 2015, 2017; Qiu et al., 2016; Huang et al., 2019). Whole-genome sequencing and *in silico* serotype determination eliminates the need for multiple serological or PCR assays, can characterize hundreds of isolates in a single assay, and differentiates between SS2 and SS1/2 and SS1 and SS14 (Athey et al., 2016).

Serotype distributions of *S. suis* vary by geographical location and provide valuable information on the prevalent subtypes affecting swine herds (Goyette-Desjardins et al., 2014). Globally, SS2 (and often including SS1/2 in the case of multiplex PCR) is the predominant serotype isolated from diseased pigs according to published studies from North America (Goyette-Desjardins et al., 2014; Gottschalk and Lacouture, 2015), Asia (China) (Goyette-Desjardins et al., 2014; Zhang et al., 2019), South America (Brazil, Argentina) (Goyette-Desjardins et al., 2014; Matajira et al., 2019), some European countries (Denmark, France, Germany, Italy, United Kingdom) (Goyette-Desjardins et al., 2014; Prüfer et al., 2019), and Australia (O'Dea et al., 2018). However, the data from Denmark, France, Italy, and the U.K. were collected prior to 2000 and may not represent current distributions in these European countries. In the Netherlands and Spain, SS9 is frequently recovered from diseased pigs, followed by SS2 and SS7 (Goyette-Desjardins et al., 2014; Willemse et al., 2016), however, an update on the distribution of serotypes in

European countries is needed to verify this shift from SS2 to SS9. The virulence of predominant serotypes was confirmed in experimental infection models in which SS1 (Smith et al., 1996; de Greeff et al., 2011), SS2 (Vecht et al., 1992; Beineke et al., 2008), SS7 (Boetner et al., 1987; Rieckmann et al., 2018), or SS9 (Beineke et al., 2008; Willemse et al., 2019) induced specific *S. suis* clinical signs and mortality with notable differences in virulence between and within serotypes.

A majority of reported human cases of *S. suis* infection are caused by SS2. The serotype distributions of human isolates in Asia and Europe, where the most cases were reported, have been heavily investigated but data from countries with sporadic cases of human infections is limited or unverified (Goyette-Desjardins et al., 2014). Overall, only serotypes 1, 2, 4, 5, 9, 14, 16, 21, 24 and 31 are currently known to cause disease in humans (Goyette-Desjardins et al., 2014; Hatrongjit et al., 2015; Kerdsin et al., 2017). Serotyping of isolates recovered from healthy pigs is of increasing importance as it provides information on the population structure of *S. suis* in carrier animals (Marois et al., 2007; Luque et al., 2010; Hoa et al., 2011; Thongkamkoon et al., 2017; Zou et al., 2018; Denich et al., 2020; Zhang et al., 2020). This is of high relevance to zoonotic transmission because carrier pigs are a major reservoir of serotypes capable of causing human infection (Hoa et al., 2011; Thongkamkoon et al., 2017; Zou et al., 2018). Serotypes 2, 9/32 (cross-reacted), and 21 were frequently isolated from tonsil samples of slaughterhouse pigs in southern Vietnam (Hoa et al., 2011). Similarly, Thongkamkoon et al. (2017) identified 8 of the 10 serotypes associated with human disease in tonsil samples of clinically healthy pigs (Thongkamkoon et al., 2017).

MLST

Multilocus sequence typing (MLST) is a nucleotide sequence-based technique for subtyping bacteria. The traditional MLST method for subtyping *S. suis* relies on the PCR amplification and sequencing of seven housekeeping genes (*aroA*, *cpn60*, *dpr*, *mutS*, *recA*, and *thrA*) (King et al., 2002). The use of a common MLST scheme affords high discrimination, reproducibility, and increased accessibility of data over the Internet. Currently there are 1654 registered sequence type (ST) profiles (as of August 2021) of *S.*

suis (pubmlst.org), however, only a small number of STs have been shown to correlate with disease.

The global distribution of *S. suis* STs also varies geographically and information on this distribution is largely limited to SS2 and SS9, which are of both animal and public health significance. ST1 (CC1) is the predominant SS2 ST isolated from diseased pigs and human cases in Europe and Asia (Goyette-Desjardins et al., 2014). ST7 (CC1) is the predominant cause of human infections in China, including the 2005 outbreak (Ye et al., 2008), while ST20 is an important cause of human infections in the Netherlands (Schultsz et al., 2012). ST25 and ST28 (both CC28, previously CC27) are the most prevalent STs among North American strains, although differences in distribution exist between Canada and the United States (King et al., 2002; Fittipaldi et al., 2011). Fittipaldi et al. (2011) determined that 95% of SS2 porcine isolates from North America belonged to STs 25 and 28, with 59% of Canadian and 75% of U.S. isolates belonging to ST25 and ST28, respectively (Fittipaldi et al., 2011). Virulence studies of different STs of SS2 demonstrated that the epidemic Chinese ST7 strain was more virulent than virulent European ST1 strains (Ye et al., 2009; Lachance et al., 2013). North American ST25 strains were intermediately virulent in a mouse infection model (Fittipaldi et al., 2011). Furthermore, Eurasian ST28 strains induced high rates of meningitis in mice while most North American ST28 strains were essentially avirulent/nonvirulent (Fittipaldi et al., 2011; Auger et al., 2016a). SS9 is strongly associated with invasive disease in European swine herds, and 89-95% of clinical SS9 strains were previously characterized as ST16 (CC16) (Schultsz et al., 2012; Willemse et al., 2019). SS7 ST29 strains were reported as an emerging *S. suis* pathotype in Europe (Rieckmann et al., 2018). ST1/CC1 and ST25, ST27 and ST28 (all CC28) have also been isolated from pig and human cases in Australia, Japan, and Thailand (Takamatsu et al., 2008b; Onishi et al., 2012; O'Dea et al., 2018).

Other standard molecular techniques

Other standard molecular techniques utilized for describing the genetic diversity of *S. suis* include pulse-field gel electrophoresis (PFGE) (Berthelot-Hérault et al., 2002;

Vela et al., 2003), randomly amplified polymorphic DNA (RAPD) (Chatellier et al., 1999; Cloutier et al., 2003), amplified fragment length polymorphism (AFLP) (Rehm et al., 2007), polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP) (Marois et al., 2006), multilocus variable number tandem repeat analysis (MLVA) (Li et al., 2010b), and ribotyping (Harel et al., 1994; Okwumabua et al., 1995). Selection of molecular epidemiological tools is influenced by reproducibility, discriminatory power, ease of use and interpretation, cost, and universal applicability (Hatrongjit et al., 2020). These techniques have provided valuable information into the epidemiology of *S. suis*. Genetic heterogeneity of *S. suis* was observed between and within swine herds with the prevalence and potential persistence of a few genotypes/clones responsible for a large proportion of clinical cases (Cloutier et al., 2003; Vela et al., 2003). Outbreak investigations, such as the analysis of isolates from the 1998 and 2005 Chinese outbreaks (Li et al., 2010b), provided molecular evidence of transmission of *S. suis*. Furthermore, the relationships between isolates from pigs of different clinical status (diseased vs healthy) (Okwumabua et al., 1995; Berthelot-Hérault et al., 2002; Cloutier et al., 2003; Marois et al., 2006; Rehm et al., 2007), host (human vs pig) (Chatellier et al., 1999; Berthelot-Hérault et al., 2002; Marois et al., 2006), and geographical region (Chatellier et al., 1999) were evaluated.

Virulence factors and markers

A growing list of putative virulence-associated factors (virulence factors and markers) has been proposed for *S. suis*. Putative virulence factors and markers include surface and secreted proteins, enzymes/proteases, and inflammatory factors that are involved at different steps of pathogenesis (see in-depth reviews (Baums and Valentin-Weigand, 2009), (Fittipaldi et al., 2012), (Feng et al., 2014), (Segura et al., 2017), (Tram et al., 2021)). Four important virulence-associated factors of *S. suis* are the capsular polysaccharide (CPS, *cps* gene loci), which is generally accepted as a virulence factor, and extracellular protein factor (EF, *epf* gene), muramidase-released protein (MRP, *mrp* gene), and suilysin (SLY, *sly* gene), which are considered "classical" virulence markers. As described above (see Serotyping section), SS2 strains are the most prevalent serotype

worldwide and are generally virulent in both pigs and humans. As a result, candidate virulence factors were mostly evaluated using SS2 (ST1 and ST7) experimental models.

The *S. suis* CPS plays an important role in the development of systemic and CNS disease, such as protecting against phagocytosis (see Pathogenesis section). However, up to 89% of nontypeable *S. suis* strains recovered from diseased pigs were reported as poorly or non-encapsulated (Bonifait et al., 2010; Gottschalk et al., 2013). In addition, non-encapsulated strains have been isolated from cases of porcine endocarditis indicating that certain non-encapsulated strains may also be virulent (Lakkitjaroen et al., 2011, 2014).

Suilysin (haemolysin) is a thiol-activated toxin and has high similarity with the pneumolysin toxin produced by the human pathogen *S. pneumoniae* (Segers et al., 1998). Invasive Eurasian strains are predominantly positive for suilysin (Segers et al., 1998; King et al., 2001). Suilysin-positive strains may promote invasion of mucosal epithelial cells and BMECs (BBB) through cytotoxic effects (see Pathogenesis section) (Tenenbaum et al., 2016). Moreover, immunization with SLY protected mice (Du et al., 2013) and partially protected pigs (Jacobs et al., 1996) against SS2 challenge. However, the lack of suilysin in various disease-associated strains, especially North American strains (Segers et al., 1998; Fittipaldi et al., 2009; Gottschalk et al., 2013), indicates suilysin should be considered an important "virulence marker" as opposed to a "virulence factor."

MRP and EF are surface/secreted proteins with variants of different molecular weights (Silva et al., 2006). SS2 strains of the MRP+/EF+ phenotype induced specific clinical signs of *S. suis* disease (lameness or CNS signs) in pigs while strains of the MRP+/EF- and MRP-/EF- phenotypes induced nonspecific or no clinicals signs, respectively (Vecht et al., 1992). Wei et al. (2009) observed similar levels of virulence (high mortality) in mice by SS2 strains with the *mrp*+/*epf*+/*sly*+ genotype (Wei et al., 2009). Invasive Eurasian strains are predominantly positive for MRP and EF (Wisselink et al., 2000; Wei et al., 2009), while invasive North American strains are predominantly MRP (or its variants)+/EF- or MRP-/EF- (Gottschalk et al., 2007; Fittipaldi et al., 2009),

although a high proportion of MRP+/EF+ strains was previously reported in the United States (Galina et al., 1996). Inactivation of the *mrp* and *epf* genes (SS1 and SS2) had no significant effect on virulence in pigs (Smith et al., 1996) but the deletion of *mrp* (SS2) reduced brain bacterial loads and histopathological signs of meningitis in infected mice (Wang et al., 2015). Together these results suggest that MRP and EF are virulence markers of *S. suis*.

While significant progress has been made in identifying *S. suis* virulence factors, the definition of "virulence" and "virulence factor" is controversial. Strains may be classified as virulent or avirulent based on the clinical condition of the animal (clinically healthy or diseased) from which the strain was isolated or based on experimental infection models (Gottschalk and Segura, 2000). Virulence may also be assigned based on the presence of virulence-associated factors. However, the criteria and results are variable across research groups. For instance, variations in experimental design (strain serotype and/or ST, route of infection), *in vitro* models (cells or blood of porcine vs mouse or human origin), and animal models (mouse vs pig) contribute to inconsistent results on virulence (Gottschalk et al., 1999; Segura et al., 2017). Intravenous inoculations reproduce systemic disease (high morbidity and mortality) but are not appropriate models for investigating virulence factor candidates involved in the initial steps of colonization and mucosal invasion (Segura et al., 2016). Intranasal *S. suis* inoculation of pigs would be preferable as a "natural" route of infection but results in no or limited disease (Pallarés et al., 2003). Several recommendations, such as use of standardized models for defining virulence and shared reference strains across research laboratories, have been made to address the problems observed with *S. suis* virulence and the discovery of *S. suis* virulence factors (Gottschalk and Segura, 2000).

Whole-genome sequencing

Advances in next generation sequencing technologies and generation of large amounts of data led to numerous programs and software tools for bacterial characterization (Carriço et al., 2013). The use of these WGS-based approaches to complement detailed clinical data has been instrumental in the development of subtyping

tools for differentiating commensal and diseased-associated (systemic and/or respiratory) strains and for the identification of virulence markers of *S. suis*. CGH analysis of 55 *S. suis* isolates demonstrated the genomic similarity of SS1 and SS2 isolates originating from Canada, the Netherlands, China, and Vietnam, and isolated from both human and pigs (Cluster A) (de Greeff et al., 2011). These isolates correlated with ST1/CC1 and the MRP+/EF+ genotype which is typical of virulent strains. On the other hand, Cluster B was mostly composed of SS7 and SS9 with low inter- and intragenic similarity and isolates with an MRP-/EF- genotype. Zheng et al. (2014) described the classification of 39 European, Chinese, and Canadian (human and porcine) isolates into three CGH groups of varying virulence: epidemic and highly virulent (E/HV), virulent (V), and intermediately or weakly virulent (I/WV) (Han Zheng et al., 2014, p. 201). The E/VH group largely consisted of SS2, ST1, and ST7, and the V and I/WV groups contained a diverse set of STs. The virulence of representative isolates was verified *in vitro* and *in vivo* (murine model).

A minimum core genome (MCG) typing system for *S. suis* classified isolates into seven MCG groups based on group-specific SNP sites (Chen et al., 2013). Zheng et al. (2014) reduced the number of single nucleotide polymorphisms (SNPs) needed to identify the seven MCG groups from 553 to 10 SNPs, providing a more practical, rapid, and cost-effective assay for use in laboratories (H. Zheng et al., 2014). Only MCG group 1 contained isolates from human infections and outbreaks, such as ST1 and ST7, and possessed the greatest number of virulence-associated genes (Chen et al., 2013). MCG groups 6 and 7 appeared to have the lowest virulence because they carried the least number of genes and a majority of the isolates from healthy pigs.

A Bayesian Analysis of Population Structure (BAPS) of 98 *S. suis* isolates from human cases and diseased pigs in the Netherlands was utilized to explore differences in zoonotic potential between isolates (Willemse et al., 2016). In the study, BAPS groups 4 (CC1) and 6 (CC20) were determined to have zoonotic potential, which was supported by a significant presence of virulence-associated genes in these isolates compared to the other 4 BAPS groups. Comparative genomic analysis of 375 human and porcine isolates

from the United Kingdom and Vietnam was able to differentiate isolates by clinical phenotype (systemic, respiratory, and non-clinical), and revealed a trend of genome-reduction (smaller genomes) but an over-representation of virulence-associated genes in systemic isolates (Weinert et al., 2015).

Long-read sequencing technology generate reads longer than 10Kb producing more contiguous genome assemblies (Amarasinghe et al., 2020). MinION nanopore sequencing of pathogens of veterinary importance is relatively new and its advantages include high mobility, low cost, and rapid turnaround. MinION sequencing was able to accurately predict the STs and antimicrobial resistance profiles of 8-10 of the *S. suis* isolates tested, demonstrating the potential use of MinION as a diagnostic tool for the characterization of pathogens (Tan et al., 2020). Bioinformatic analyses have also aided in the identification of candidate vaccine proteins against *S. suis* for subsequent evaluation in vaccine trials. For example, bioinformatics analyses can be applied to determining origin and conservation of candidate genes or proteins (Brockmeier et al., 2018; Feng et al., 2018), determining amino acid sequence and tertiary structure (Zhang et al., 2014), and predicting subcellular location and identifying protective antigen sequences (Geng et al., 2008; Liu et al., 2009, p. 200).

1.3 Tables

Table 1.1 Worldwide distribution of major *S. suis* serotypes and sequence types (STs) involved in swine clinical cases

Country		Serotypes	STs	References
North America				
	Canada	1/2, 2-4, 7, 8	25, 28 (SS2)	(Goyette-Desjardins et al., 2014; Gottschalk and Lacouture, 2015)
	United States	1/2, 2, 3, 6-8	25, 28 (SS2)	(Fittipaldi et al., 2009; Goyette-Desjardins et al., 2014)
South America				
	Brazil	1-3, 1/2, 7, 14		(Goyette-Desjardins et al., 2014; Calderaro et al., 2016; Matajira et al., 2019)
	Argentina	2	1 (SS2)	(Callejo et al., 2016; Doto et al., 2016)
Europe				
	Germany	1, 2, 1/2, 7, 9, 14	1 (SS2) 29 (SS7)	(Wisselink et al., 2000; Goyette-Desjardins et al., 2014; Prüfer et al., 2019)
	The Netherlands	2, 7, 9	1, 20 (SS2) 16 (SS9)	(Goyette-Desjardins et al., 2014; Willemse et al., 2019)
	Spain	2, 7, 9	1 (SS2) 123, 125 (SS9)	(Goyette-Desjardins et al., 2014; Zheng et al., 2018)

	Switzerland	6, 9	13 (SS1) 1104 (SS6) 1105 (SS9)	(Scherrer et al., 2020)
	Belgium	1, 2, 9, 14		(Wisselink et al., 2000)
	Denmark	2, 7		(Wisselink et al., 2000, p. 200)
	France	1-3, 7, 9	1 (SS2)	(Wisselink et al., 2000; Goyette-Desjardins et al., 2014)
	Italy	2, 1/2, 9		(Wisselink et al., 2000)
	United Kingdom	2	1 (SS2, SS14) 29 (SS7)	(Wisselink et al., 2000; Goyette-Desjardins et al., 2014)
	Sweden	5, 11, 15, NT		(Werinder et al., 2020)
Asia				
	China	2, 3, 4, 7, 9	1, 7, 28, 117 (SS2) 29, 117, 242, 243	(Goyette-Desjardins et al., 2014; Zhang et al., 2019; Peng et al., 2020)
	South Korea	2, 1/2, 3, 8, 9		(Goyette-Desjardins et al., 2014; Gurung et al., 2015; Oh et al., 2017)
	Japan	2	1, 28 (SS2)	(Goyette-Desjardins et al., 2014)
Australia		2, 1/2, 3	1 (SS1/2) 25, 28 (SS2) 27 (SS3)	(O'Dea et al., 2018)

Table 1.2 Potential markers of virulence for major *S. suis* serotypes and STs

Serotypes	Potential markers of virulence	References
2	mrp, epf, sly: epf+/mrp+/sly+ (ST1), epf-/mrp-/sly- (ST25), epf-/mrp*/sly+/- (ST28)	(Silva et al., 2006; Li et al., 2010b)
	pilus clusters: srtBCD+/srtF+ (ST1), srtF-/srtG+ (ST25), srtF+/srtG+ (STs 27, 28)	(Takamatsu et al., 2009; Fittipaldi et al., 2011; Onishi et al., 2012)
	ofs: type-1 ofs (CC1), type-3 and 4 ofs (CC27)	(Takamatsu et al., 2008a)
	rgg, endoD, SMU_61-like and SpyM3_0908	(Dong et al., 2015)
9	mrp*/epf−/sly+	(Zheng et al., 2018)
	cps9, cpsK, fbps, manN, gapdh, virB4, virD4, sao, revS, SSU0207, SSU1589, SSUST30534	(Dong et al., 2017; Zheng et al., 2018; Willemse et al., 2019; Scherrer et al., 2020)
other serotypes	SSU0207, SSU1589, SSUST30534	(Wileman et al., 2019)

CHAPTER 2

Serotype and genotype (Multilocus Sequence Type) of *Streptococcus suis* isolates from the United States serve as predictors of pathotype

April A. Estrada,[a] Marcelo Gottschalk,[b] Stephanie Rossow,[c] Aaron Rendahl,[a] Connie Gebhart,[a,c] Douglas G. Marthaler[c,d]

[a]Department of Veterinary and Biomedical Sciences, College of Veterinary Medicine, University of Minnesota, St. Paul, Minnesota, USA

[b]Faculty of Veterinary Medicine, University of Montreal, Saint-Hyacinthe, Quebec, Canada

[c]Veterinary Diagnostic Laboratory, College of Veterinary Medicine, University of Minnesota, St. Paul, Minnesota, USA

[d]Veterinary Diagnostic Laboratory, College of Veterinary Medicine, Kansas State University, Manhattan, Kansas, USA

Estrada AA et al. Serotype and genotype (Multilocus Sequence Type) of *Streptococcus suis* isolates from the United States serve as predictors of pathotype. *J Clin Microbiol*. 2019;57(9):e00377-19. doi:10.1128/JCM.00377-19

2.1 Synopsis

Streptococcus suis is a significant cause of mortality in piglets and growing pigs worldwide. The species contains pathogenic and commensal strains, with pathogenic strains causing meningitis, arthritis, endocarditis, polyserositis, and septicemia. Serotyping and multilocus sequence typing (MLST) are primary methods to differentiate strains, but the information is limited for strains found in the United States. The objective of this study was to characterize the diversity of 208 *S. suis* isolates collected between 2014 and 2017 across North America (mainly the United States) by serotyping and MLST and to investigate associations between subtype and pathotype classifications (pathogenic, possibly opportunistic, and commensal), based on clinical information and site of isolation. Twenty serotypes were identified, and the predominant serotypes were 1/2 and 7. Fifty-eight sequence types (STs) were identified, and the predominant ST was ST28. Associations among serotypes, STs, and pathotypes were investigated using odds ratio and clustering analyses. Evaluation of serotype and ST with pathotype identified a majority of isolates of serotypes 1, 1/2, 2, 7, 14, and 23 and ST1, ST13, ST25, ST28, ST29, ST94, ST108, ST117, ST225, ST373, ST961, and ST977 as associated with the pathogenic pathotype. Serotypes 21 and 31, ST750, and ST821 were associated with the commensal pathotype, which is composed of isolates from farms with no known history of *S. suis*-associated disease. Our study demonstrates the use of serotyping and MLST to differentiate pathogenic from commensal isolates and establish links between pathotype and subtype, thus increasing the knowledge about *S. suis* strains circulating in the United States.

Keywords MLST, *Streptococcus suis*, multilocus sequence typing, pathogenic, pathotype, porcine, serotyping

2.2 Introduction

Disease caused by *Streptococcus suis* is a significant economic and welfare concern in the swine industry. *S. suis* is a Gram-positive bacterium, and the species contains pathogenic and commensal strains. Pathogenic *S. suis* strains are associated with meningitis, arthritis, endocarditis, polyserositis, and septicemia in piglets and growing pigs (Clifton-Hadley, 1984; Reams et al., 1994), and *S. suis* strains isolated from neurological or systemic tissues (brain/ meninges, joints, and heart) are commonly considered the primary pathogens (Galina et al., 1992; Reams et al., 1994; Weinert et al., 2015). Commensal strains normally reside in the upper respiratory tract of pigs, with pigs commonly serving as carriers (Clifton-Hadley and Alexander, 1980; Arends et al., 1984; Clifton-Hadley, 1984). *S. suis* can be an opportunistic pathogen associated with coinfections with other bacterial and viral pathogens (Galina et al., 1992; Reams et al., 1994). In addition, some *S. suis* strains have zoonotic potential, causing meningitis in humans (Gottschalk et al., 2007).

Serotyping and multilocus sequence typing (MLST) are commonly used to differentiate *S. suis* strains, and numerous subtypes exist within the *S. suis* species. Traditionally, serotyping uses the coagglutination test with reference antisera to subtype *S. suis* strains, and 35 known *S. suis* serotypes (1 to 34 and 1/2) exist. Six of those serotypes have been reclassified as *Streptococcus orisratti* (serotypes 32 and 34), *Streptococcus parasuis* (serotypes 20, 22, and 26), or *Streptococcus ruminantium* (serotype 33) (Hill et al., 2005; Tien et al., 2013; Tohya et al., 2017). The development of PCR-based and whole-genome sequencing (WGS) techniques can also be used to serotype and identify strains that were previously nontypeable through traditional serological methods (Kerdsin et al., 2014; Athey et al., 2016).

A 1992 United States study investigated the serotype distribution of *S. suis* in porcine samples from Minnesota and reported the prevalence of serotypes 2 to 9 and 11, of which serotype 2 was the predominant serotype associated with neurological disease (Galina et al., 1992). A 1993 U.S. study identified serotypes 1 to 8 and 1/2 in naturally infected pigs primarily from a single state, with serotype 2 being the predominant

serotype, followed by serotypes 3, 4, 7, 8, 1, 5, 1/2, and 6 (Reams et al., 1993). A large U.S. study in 2009 investigated the serotype distribution of *S. suis* strains collected from 2003 to 2005 from 17 states, illustrating that the distribution of strains was similar to Canada (Fittipaldi et al., 2009). In both countries, serotypes 1/2, 2, 3, 7, and 8 were most prevalent in diseased pigs (Messier et al., 2008; Fittipaldi et al., 2009) which is dissimilar to the distribution in Europe, in which serotype 2 occurs at a considerably higher percentage of isolates than in North America (Wisselink et al., 2000).

MLST is a nucleotide sequence-based technique for subtyping bacteria, and a standard MLST scheme has been developed for *S. suis*, with 1,161 registered sequence type (ST) profiles as of 28 February 2019 (King et al., 2002) (pubmlst.org). Global MLST studies of *S. suis* identified ST1, ST25, and ST28 as the most prevalent STs in swine (de Greeff et al., 2011; Onishi et al., 2012; Schultsz et al., 2012; Zhu et al., 2013). In North America, ST25 and ST28 are more common among strains recovered from diseased animals, while ST1 strains are more prevalent in Europe and Asia (de Greeff et al., 2011; Fittipaldi et al., 2011; Schultsz et al., 2012). However, these studies address MLST for serotype 2 strains and may not apply to the remaining serotypes.

Previously, studies have classified isolates into pathotypes based on clinical information and site of isolation (Galina et al., 1992; Weinert et al., 2015). Our objective was to combine information on pathotype with serotype and ST to address the limited information on current *S. suis* strains circulating within the United States. In total, 208 porcine *S. suis* isolates from North America were characterized by serotyping and MLST to determine the population and distribution of *S. suis* in the United States. Furthermore, the serotype and MLST data were used to investigate associations with the pathogenic and commensal pathotypes with the goal to identify pathogenic- and commensal-specific serotype and MLST patterns. Identifying the major disease-causing strains can promote the development of treatment and control plans. Our research seeks to identify pathogenic strains to track isolates in an outbreak, select strains for a vaccine, and develop effective treatment and control plans.

2.3 Materials and methods

2.3.1 Selection of *S. suis* isolates

A total of 208 *S. suis* isolates were selected for the project. Most of the *S. suis* isolates were obtained from routine diagnostic cases submitted between April 2014 and July 2017 to the University of Minnesota Veterinary Diagnostic Laboratory (UMNVDL) or the Kansas State Veterinary Diagnostic Lab (KSVDL). Further commensal isolates were collected from 9 different farms with a lack of systemic *S. suis* clinical disease. Isolates that met our pathotype criteria (defined below) were selected from as many states as possible ($n = 20$) to minimize sample bias and increase geographic diversity to represent the major regions of the U.S. swine industry. *S. suis* isolates were verified to the species level by matrix-assisted laser desorption ionization–time of flight mass spectrometry (MALDI-TOF MS) (Microflex device, Bruker Daltonics GmbH, Germany) (Pérez-Sancho et al., 2017).

Multiple isolates may be recovered from healthy pigs due to the native microflora of the upper respiratory tract, while a single isolate is generally responsible for systemic infections (Staats et al., 1997). To limit the bias in isolating and selecting strains associated with clinical signs, a pathotype category system was developed for the *S. suis* isolates similar to previously published methods (Higgins et al., 1992; Weinert et al., 2015). “Pathogenic” isolates were obtained from the brain/meninges, joint, heart, or liver and reported as the primary cause of meningitis, arthritis, epicarditis, or septicemia in diagnostic reports by pathologists. “Possibly opportunistic” isolates were from lung samples submitted to the diagnostic lab from pigs without signs of neurological or systemic disease and included two isolates from nasal samples from farms with a clinical outbreak of *S. suis* disease. “Commensal” isolates were from laryngeal, tonsil, or nasal samples retrieved from farms with no known history or current control methods for *S. suis* disease.

2.3.2 Serotyping, MLST via whole-genome sequencing

Isolates were recultured for 24 to 48 h at 37°C on blood agar plates (tryptic soy agar [TSA] with 5% sheep blood) (Thermo Fisher Scientific, Waltham, MA, USA) and sent for serotyping to the bacterial serology laboratory at the Diagnostic Service of the Faculty of Veterinary Medicine of the Université de Montréal, Canada. The serotyping was done through the coagglutination test with reference antisera (Gottschalk et al., 1989; Marcelo Gottschalk et al., 1991; M. Gottschalk et al., 1991; Higgins et al., 1995). Nontypeable samples (samples which failed to react with the serum panel, autoagglutinated, or reacted to several sera) were further serotyped by PCR (Okura et al., 2014), a technique that cannot differentiate serotype 2 from 1/2 and serotype 1 from 14.

The *S. suis* DNA was extracted using the protocol for cultured cells from the QIAamp DNA kit (Qiagen Inc., Germantown, MD, USA) and submitted to the University of Minnesota Genomic Center (UMGC, St. Paul, MN, USA) for library preparation using Nexture TX (Illumina, San Diego, CA), and next-generation sequencing was performed on a HiSeq 2500 instrument (Illumina) with 250-bp paired-end reads. Illumina sequencing reads for each isolate were processed using Trimmomatic (Bolger et al., 2014) with an average quality cutoff of 20 (2.3 million average reads per sample). Strains were again confirmed as *S. suis* by having a 96.6% to 100% nucleotide identity to the 1,662-bp *S. suis*-specific recombination/repair protein (*recN*) sequence (*Streptococcus suis* 05HAS68, GenBank accession number CP002007) using the *S. suis* serotyping pipeline (Ishida et al., 2014).

The isolates with serotypes 2 or 1/2, 1 or 14, or lacking a subtype using traditional serotyping or PCR methods were serotyped using the WGS data and the *S. suis* serotyping pipeline (https://github.com/ streplab/SsuisSerotyping_pipeline) (Athey et al., 2016). The pipeline uses the *cpsK* gene (*Streptococcus suis* NSUI002, GenBank accession number CP011419) missense mutation at position 161 to differentiate serotypes 2 and 1/2 and serotypes 1 and 14. Isolates undifferentiable by WGS were categorized as serotype “1or14” or NT (nontypeable) in the downstream analysis.

In silico MLST analysis was performed using the Short Read Sequence Typing for Bacterial Pathogens (SRST2) program (http://katholt.github.io/srst2), which maps

reads to MLST references (Inouye et al., 2014). The ST allele sequences and profiles were obtained from the *S. suis* MLST database (https://pubmlst.org/ssuis/) (Jolley and Maiden, 2010). Novel ST allele sequences were confirmed by PCR amplification and Sanger sequencing of the *aroA*, *cpn60*, *dpr*, *gki*, *mutS*, *recA*, or *thrA* genes (King et al., 2002). The primers used for the amplification and sequencing of the *mutS* gene were mutS forward (5'-AAGCAGGCAGTCGGCGTGGT-3') and mutS reverse (5'-AGTACAAACTACCATGCTTC-3') as described (Rehm et al., 2007). STs were grouped into major clonal complexes (CCs) using the entire MLST database and the eBURST software (Feil et al., 2004). Groups were defined with the strict parameters for determining single-locus variants (match of 6 or more loci). The entire *S. suis* MLST database was displayed as a single eBURST diagram by setting the group definition to zero of seven shared alleles.

2.3.3 MLST clustering analysis

Alignments, sequence identity calculations, and construction of the MLST sequence identity heatmap for basic clustering analysis were performed with R software (v.3.4.3) (R Core Team, 2017) and R packages (Charif and Lobry, 2007; Day, 2012; Bodenhofer et al., 2015; Warnes et al., 2016). The concatenated sequences of the seven MLST alleles were aligned with MUSCLE (v.3.8.31) (Edgar, 2004), and sequence identities were calculated. The sequence identity scores were used to generate a heatmap based on Euclidian distances and neighbor joining clustering.

2.3.4 Statistical analysis

Basic data transformation and plotting for statistical analyses were performed using R software and R packages (Wickham, 2009, 2017; Chen et al., 2018). Ternary plots of subtypes and pathotypes were generated using the R package Ternary (v.1.0.2) (Smith, 2017). The pathotype boundaries were assigned and color-coded using 50% as a cutoff. Odds ratio (OR) analysis was used to test all pathotype-subtype combinations containing more than a single isolate, and 95% confidence intervals (CIs) were generated using Fisher's exact test. For each combination, the 2 by 2 table was created comparing

that pathotype and subtype against all others. Similar 2 by 2 tables were generated for testing pathotype and serotype-ST-combinations by chi-square and Fisher's exact tests. ORs greater than 1 with a 0.3 minimum lower limit were considered biologically significant. The minimum lower limit of 0.3 was calculated as the average lower limit among the combinations, is specific to our data set, and was selected for the identification of biologically meaningful relationships. An infinite (Inf) OR for a pathotype-subtype combination refers to a subtype that occurred in only one pathotype. The associations within and between types were investigated using multiple correspondence analysis (MCA), with the FactoMineR (v.1.41) and factoextra (v.1.0.5) packages (Lê et al., 2008; Kassambara and Mundt, 2017), by setting the serotype, ST, and pathotype as the three variables.

2.3.5 Data availability

The reads associated with the samples were deposited in the NCBI Sequence Read Archive under accession numbers SRR9123061 to SRR9123268 (see Table S2.1).

2.4 Results

2.4.1 Serotype and ST distributions of *S. suis* in the United States

Characterization of S. suis isolates by serotyping and MLST

A total of 208 *S. suis* isolates were characterized, of which 203 were from the United States, 4 from Canada, and 1 from Mexico (Fig. 2.1). The clinical history and tissue of origin of the isolates were used to determine the pathotype, and the 208 isolates were classified as pathogenic (n = 139), possibly opportunistic (n = 47), and commensal (n = 22) (Table 2.1). The *recN* segment from *S. suis* was identified in the whole-genome sequences of all the 208 strains (>99% coverage of the gene and 40× to 314× depth), indicating that the isolates were *S. suis*.

Serotyping identified 20 different serotypes representing 94.2% of the strains, while 5.3% were nontypeable (indicated as NT) and 0.5% could not be differentiated

between serotype 1 and 14 (n = 1) by coagglutination, PCR, or WGS (see Table S2.2 in the supplemental material). The predominant serotypes were 1/2 (n = 54) and 7 (n = 23).

In silico MLST analyses were performed on the WGS data, and the samples had an average depth of 155× across the seven loci. STs could not be determined for four isolates because one housekeeping gene necessary for MLST classification was not identified in these isolates (referred to as NF, see Table S2.1 in the supplemental material). Fifty-eight different STs were identified for the remaining 204 isolates, indicating high diversity among the isolates (see Table S2.3 in the supplemental material). Twenty of these STs were previously defined, while 38 were newly identified (961 to 969, 971 to 998, and 1001; n = 56). The predominant ST was ST28 (n = 52), followed by ST94 (n = 18), ST1 and ST108 (n = 17 each).

Relationship between serotypes and STs

The distribution of STs by serotype illustrated the diversity of the *S. suis* strains (Fig. 2.2). Fifteen of the 20 serotypes identified contained multiple STs, with the number of different STs within a single serotype ranging from 2 to 8. The predominant serotype 1/2 contained three STs (ST28 [n = 44], ST961 [n = 8], and ST982 [n = 1]). Serotypes 8, 14, 24, 28, and 29 contained a single ST each, namely, ST87, ST1, ST94, ST968, and ST972, respectively. However, serotypes 24, 28, 29, and 1or14 contained only a single isolate.

Distribution of pathotypes by serotype and ST

The distribution of *S. suis* pathotypes by serotype is shown in Fig. 2.3. Serotype 1/2 contained the most pathogenic isolates (n = 45/54), followed by serotypes 7 (n = 19/23), 2 (n = 14/17), and 1 (n = 11/ 11). Serotypes 1 and serotype 14 (n = 5) were composed entirely of pathogenic isolates. The distribution of pathotypes by ST is shown in Fig. 2.4. ST28 contained the most pathogenic isolates (n = 42/52), followed by ST1 (n = 17/17), ST94 (n = 14/18), and ST108 (n = 14/17), of which only ST1 was composed entirely of pathogenic isolates. Additional pathogenic STs included ST961 (n = 9/10),

ST977 (*n* = 7/9), and ST13 (*n* = 5/ 5). Twelve STs contained isolates only classified as commensal, with ST750 and ST821 containing more than a single commensal isolate.

2.4.2 Associations among pathotypes, serotypes, and STs by analysis of proportions and OR

Associations between pathotype and serotype

Proportions and OR analyses were used to investigate pathotype associations with serotype for serotypes (proportions) or serotype-pathotype combinations (OR analysis) that contained more than one isolate. Between 80% and 100% of serotypes 1, 1/2, 2, 7, 14, and 23 were classified as the pathogenic pathotype (Fig. 2.5A), and these associations were supported by OR analysis (Fig. 2.5B). In the ternary plot, serotypes 3, 5, and 9 demonstrated a moderate association with the pathogenic pathotype, with 56% to 63% of isolates classified as pathogenic. However, the association between pathotype and serotype was not supported by OR analysis. OR analysis supported associations of serotypes 10 and 12 with the possibly opportunistic pathotype, with 67% of isolates classified as possibly opportunistic in the ternary plot. Serotypes 21 and 31, with 67% to 80% of isolates classified as commensal in the ternary plot, were supported as commensal pathotypes by OR analysis.

Associations between pathotype and ST

Proportions and OR analysis were used to investigate pathotype associations with ST for STs (proportions) or ST-pathotype combinations (OR analysis) that contained more than one isolate. The ternary plot of the 58 STs (and the NF category) illustrated a clear differentiation by pathotype for all STs except ST87 and ST119 (approximately 50% pathogenic) (Fig. 2.6A). Twelve STs and the NF category contained over 75% of isolates classified as pathogenic, including ST1, ST13, ST25, ST28, ST29, ST94, ST108, ST117, ST225, ST373, ST961, and ST977, which demonstrated the same associations by OR (Fig. 2.6B). ST969 had an association with the possibly opportunistic pathotype, which was supported by OR. The commensal pathotype demonstrated a strong association with ST750 and ST821, which was supported by OR analysis.

Odds ratio and MCA of pathotypes, serotypes, and STs

Initially, OR was used to investigate the relationships between pathotype and serotype-ST-combinations, but significance relationships were lacking for the combinations (OR data not shown). Then, MCA was performed to analyze the possible relationships among all serotypes, STs, and pathotypes (Fig. 2.7). The first and second dimensions of the analysis only represent 6% of the data. The ellipses represent 95% of isolates in each pathotype. All the subtypes demonstrating a strong association with the pathogenic pathotype by OR analysis (Fig. 2.5 and 2.6) fell within the overlapping 95% ellipses for multiple pathotypes by MCA (Fig. 2.7). Five serotypes and 13 STs in the commensal pathotype lacked overlapping ellipses. Serotypes 21 and 31 lacked any isolates with the pathogenic pathotype (Fig. 2.3), while ST750 and ST821 contained only isolates with the commensal pathotype (Fig. 2.4). The limited representations of the MCA data (6% variance) and the overlapping ellipses indicate a lack of relationship between serotype, ST, and pathotype, highlighting potential confounding factors for predicting pathogenic isolates based on both serotyping and MLST together. Thus, the relationship between pathotype, serotype, and ST is lacking for the pathogenic and possibly opportunistic pathotypes.

2.4.3 Associations between pathotype and MLST CC by analysis of proportions and OR

Identification of S. suis CCs

To investigate the population structure of our *S. suis* isolates by MLST, the STs were assigned into CCs defined by eBURST, using the entire *S. suis* MLST database and our 58 STs (Fig. 2.8 and Table S2.3). Using the stringent definition (six of seven shared alleles) for defining a CC, five CCs (CC1, CC28, CC94, CC104, and CC750) with a primary founder were identified from our set of STs. However, multiple STs (n = 30) did not form a CC or formed a CC without a primary founder (Table 2.2). The most diverse CC (CC94) contained isolates from 13 of the 28 STs assigned into CC, compared with

CC1, CC28, CC104, and CC750, which contained isolates from 4, 7, 1, and 3 STs, respectively.

Associations between pathotype and CC

Patterns between CC and pathotype were investigated by proportions and OR analysis. CC1, CC28, CC94, and CC104 were associated with the pathogenic pathotype, and the association was supported by OR analysis (Fig. 2.9). CC750 was associated with the commensal pathotype and was supported by OR analysis, with 83% of isolates classified as the commensal pathotype. The STs among the group of isolates lacking a CC did not associate with any pathotype.

Associations among pathotypes, serotypes, and MLST by clustering analysis

To investigate genetic relationships among the samples and possible associations among serotype, genotype, and pathotype classifications, the MLST allelic sequences were clustered, illustrating 75% to 100% nucleotide sequence identity (Fig. 2.10). The five main CCs were identifiable in the MLST sequence identity heatmap as clusters of genetically similar strains. CC28 contained the most isolates with the pathogenic pathotype (*n* = 58/70) (Table 2.2) and was mostly composed of ST28 (predominant ST, *n* = 52) belonging to serotype 1/2 (predominant serotype, *n* = 44) and serotype 2 (*n* = 8) (Fig. 2.10). CC28 also contained ST961 (*n* = 10) belonging to serotype 1/2 (*n* = 8) and serotype 2 (*n* = 2). MLST clustering analysis demonstrated clustering of CC28 and CC104, and the latter consisted of only three isolates with the pathogenic pathotype (serotype 7, ST225).

CC1 was divided into two groups and clustered with CC750 and isolates without a CC. The first cluster of CC1 contained a concentration of isolates in the pathogenic pathotype (*n* = 17/28), while the second cluster contained 4 pathogenic isolates, 6 possibly opportunistic isolates, and a single isolate with the commensal pathotype (Table 2.2). The isolates within the first cluster of CC1 were predominantly characterized as ST1 (serotype 1 [*n* = 7/17], serotype 2 [*n* = 5/17], and serotype 14 [*n* = 5/17]). Lacking a CC, the ST13 isolates (*n* = 5; serotype 1 [*n* = 4] and serotype 1or14 [*n* = 1]) clustered with

CC1 isolates, demonstrating a possible genetic relatedness to isolates of CC1 and the pathogenic pathotype. Serotypes 1, 2, and 14 and ST1 and ST13 were also associated with isolates of the pathogenic pathotype by proportions and OR. Inversely, CC750 (*n* = 6) consisted of isolates with the commensal (*n* = 5) and possibly opportunistic (*n* = 1) pathotypes and was predominantly composed of isolates characterized as nontypeable (*n* = 5/6) and ST750 (*n* = 4/6). Interestingly, CC750 was closely related to the group of isolates lacking a CC (*n* = 31), which consisted of isolates with the commensal pathotype (*n* = 12/31, multiple serotypes and novel STs), providing further evidence for the association between CC750 and the commensal pathotype.

CC94 was predominantly composed of isolates with the pathogenic pathotype (*n* = 43/60) but contained isolates with all three pathotypes (Table 2.2). The isolates within CC94 with the pathogenic pathotype were predominantly characterized as ST94 (*n* = 14/43; serotype 7 [*n* = 10/14]), ST108 (*n* = 14/43; serotype 23 [*n* = 8/14]), and ST977 (*n* = 7/43; serotype 5 [*n* = 6/7]). Serotype 7 was the second-most predominant serotype and was associated with the pathogenic pathotype by proportions and OR. Clustering analysis identified CC1, CC28, CC104, and a subset of CC94 as corresponding to the pathogenic pathotype. Isolates within these CCs were predominantly characterized as serotypes 1, 1/2, 2, 7, 14, and 23 and ST1, ST13, ST28, ST94, ST108, ST961, and ST977, providing further evidence of these subtypes corresponding to the pathogenic pathotype.

Supplementary material

Supplemental material for this chapter may be found at doi 10.1128/JCM.00377-19.

2.5 Discussion

S. suis is an important swine pathogen, often resulting in neurological and systemic disease caused by pathogenic strains. However, much is still unknown about the population structure of *S. suis* in the United States. In this study, we utilized serological and molecular typing techniques to investigate the serotype and ST distributions of U.S. isolates. Fourteen of the 20 *S. suis* serotypes identified in this study were recovered from pigs with clinical disease (*n* = 139). The predominant pathogenic serotypes identified in

this study were 1/2 (*n* = 45), 7 (*n* = 19), and 2 (*n* = 14), which have been previously identified as the predominant serotypes from diseased pigs in North America (Messier et al., 2008; Fittipaldi et al., 2009; Gottschalk et al., 2013; Goyette-Desjardins et al., 2014). While serotypes 2 and 3 are considered predominant pathogenic serotypes in North America, only 10.6% of the strains in our study were recovered from diseased pigs. Furthermore, the serotype distribution from our study differed from European studies, in which serotypes 2 and 9 are predominant (Vela et al., 2003; Goyette-Desjardins et al., 2014). The higher prevalence of serotype 1/2 in North America could be due to a common evolutionary lineage with serotype 2. Genetic analysis by PCR-based serotyping of the *cps* loci demonstrated serotypes 1/2 and 2 share the same genetic profile and cannot be differentiated by serotype-specific *cps* loci (Kerdsin et al., 2014; Athey et al., 2016). Sequencing of the *cpsK* gene reveals a missense mutation permitting the differentiation of serotypes 2 and 1/2 (Athey et al., 2016), but a PCR protocol has not been implemented yet to differentiate these serotypes.

In our study, the geographic distribution of *S. suis* was from 20 different states (Table S2.1), which represent the major swine-producing states in the United States. Variability in the serotype distribution of *S. suis* has been reported within the same country, which is likely due to natural differences in geographic distribution (Reams et al., 1993). Geographic distribution of the *S. suis* serotypes in our study identified serotype 1/2 in 13 of the 20 states, with a concentration in 5 of the 20 states, possibly displaying a geographic distribution pattern of serotype 1/2 in the United States. Serotype 1/2 is also a frequent serotype found in Canada, although at lower levels than serotypes 2 and 3 (Gottschalk and Lacouture, 2015). This prevalence of serotype 1/2 in Canada may contribute to the U.S. serotype distribution through the transport of pigs between the two countries (Goyette-Desjardins et al., 2014). Transport of livestock has been associated with geographic invasion or the emergence of a pathogen in a novel geographic area (Pasma et al., 2008; Nelson et al., 2011; Lowe et al., 2014). While most transport of pigs to the United States head to harvest facilities, new breeding stock of pigs could be colonialized with new *S. suis* strains, which could result in the spread of new strains to

downstream swine farms. Whole-genome analysis of the U.S. and Canadian serotype 1/2 strains would further clarify the relationship between U.S. and Canadian 1/2 strains.

We anticipated identifying a large number of novel ST profiles due to the inclusion of commensal and possibly opportunistic samples, which are not generally subjected to subtyping by MLST. As a result of this study, 38 novel ST profiles were submitted to the *S. suis* MLST database. Of the 58 STs identified here, 24 STs were isolated from pigs with clinical disease, and the predominant STs were ST28 ($n = 42$), followed by ST1 ($n = 17$), ST94 ($n = 14$), and ST108 ($n = 14$). In a previous Canadian study in 2011, ST25 was the predominant ST found in Canada, while ST28 was the predominant ST found in the United States (Fittipaldi et al., 2011). Our results confirm ST28 as a predominant pathogenic pathotype, while ST25 represents only 1% of the strains recovered from diseased pigs ($n = 2$). The reason for this low percentage of ST25 isolates in the United States is unclear, and updated ST analysis of *S. suis* strains from Canada is needed to confirm ST25 as the predominant ST in that country. Our ST distribution also differs from that of European and Asian countries in which ST1 strains, largely characterized as serotype 2, are predominant in diseased pigs (Goyette-Desjardins et al., 2014; Auger et al., 2016b).

Proportions, OR, and clustering analysis illustrated potential relationships among pathotypes, serotypes, and STs. While multiple pathogenic serotypes and STs were identified in our study, this discussion focuses on serotype and STs with more than four isolates in the pathogenic pathotype. Serotypes 1, 1/2, 2, 7, 14, and 23 as well as ST1, ST13, ST28, ST94, ST108, ST961, and ST977 were frequently identified as pathogenic strains. Based on our pathotype classifications, isolates characterized as pathogenic were linked to neurological or systemic disease, and our analyses provide evidence that these subtypes are potential indicators of virulence. As discussed previously, serotypes 2 and 1/2 are predominant serotypes identified from diseased pigs in North America, supporting our observations of these serotypes as pathogenic strains by proportions, OR, and clustering analysis (Messier et al., 2008; Fittipaldi et al., 2009; Gottschalk et al., 2013; Goyette-Desjardins et al., 2014; Gottschalk and Lacouture, 2015).

Serotypes 1 and 7 are more prevalent in diseased pigs in some European countries than in North America, and pathogenic serotype 1 strains have been linked to the production of muramidase-released protein (MRP), extracellular-factor protein (EF), and suilysin (SLY). Pathogenic serotype 1 strains have been characterized as producing both MRP and EF, with variable production of SLY (Wisselink et al., 2000; de Greeff et al., 2011). In one study (de Greeff et al., 2011), four of the six serotype 1 strains were MRP+EF+SLY+ and five of the six were either ST1 or ST13, indicating a correlation between serotype 1, ST1, ST13, and virulence. Interestingly, the serotype 1 isolates in the current study were either ST1 (n = 7/11) or ST13 (n = 4/11) and were associated with the pathogenic pathotype, supporting the previous study. Serotype 7 was the second-most common serotype identified in this study, and 19/23 isolates were characterized as the pathogenic pathotype. Virulence studies on serotype 7 strains demonstrating clinical disease in pigs are limited, but a previous *in vivo* study associated serotype 7 with septicemia and arthritis, with rare cases of meningitis (Boetner et al., 1987). These findings support the classification of serotype 7 as pathogenic.

This study demonstrates that ST appears to be a stronger predictor of pathotype than serotype. While experimental mouse models have demonstrated the virulence of serotype 2 ST1, ST25, and ST28 (Fittipaldi et al., 2011; Auger et al., 2016b), our analyses also illustrated ST1, ST13, ST28, ST94, ST108, ST961, and ST977 (of various serotypes) as pathogenic. As mentioned previously, we hypothesize that Canadian and U.S. serotype 2 and serotype 1/2 strains share a evolutionary lineage. If so, the observed virulence of serotype 2 ST28 in previous studies may support the virulence of serotype 1/2 ST28, as predicted in our study. Whole-genome single nucleotide polymorphism (SNP)-based phylogenetic analysis *S. suis* serotype 2 ST28 strains revealed a unique clade composed of virulent strains capable of inducing severe disease in a murine infection model (Athey et al., 2015). These strains demonstrated differences in virulence to reference serotype 2 ST28 strains of low virulence. Recently, a study characterized pathogenic Australian serotype 1/2 ST1 strains by core genome single nucleotide polymorphisms and linked the genetic similarity to pathogenic serotype 1/2 ST1 strains from the United Kingdom and Vietnam (O'Dea et al., 2018). Our clustering analysis

indicates that ST1, ST13, ST94, ST108, ST961, and ST977 may also be pathogenic. It would be of interest to further investigate the virulence properties of serotype 1/2 ST28, as well as ST1, ST13, ST94, ST108, ST961, and ST977 strains isolated in the United States.

In addition to strains in CC1, CC28, and CC104, serotype 9 strains belonging to CC16 (previously CC87) have been isolated from pigs with invasive disease (Schultsz et al., 2012). However, the low percentage of serotype 9 strains in our study is reasonable because serotype 9 is predominant in diseased pigs from the Netherlands (Wisselink et al., 2000). The serotype 9 strains in this study belong to multiple CCs or occur as singletons and did not demonstrate associations with pathotype. Serotype 9 isolates from diseased and healthy pigs in China were characterized into multiple STs and demonstrated high diversity among the isolates (Dong et al., 2017). The majority of these serotype 9 isolate STs occurred as singletons and did not form major clonal complexes.

Inversely, commensal *S. suis* serotypes 21 and 31 and ST750 and ST821 were identified by proportions, OR, and cluster analysis. Studies on *S. suis* from North America have observed a prevalence of serotype 21 from healthy pigs (Gottschalk et al., 1989; Marcelo Gottschalk et al., 1991). However, previous studies have identified a limited number of serotype 31 strains from pigs with typical clinical signs of *S. suis* disease (Gottschalk et al., 2013; Gottschalk and Lacouture, 2015; Wongsawan et al., 2015; Prüfer et al., 2019). The association between serotype 31 and pathotype remains unclear and requires further investigation.

Associations among serotypes, STs, and pathotypes, although identified by individual analyses, were not evident in the MCA, indicating both serotype and ST together could not indicate pathotype. We investigated additional approaches, such as chi-square and Fisher's exact tests, but these tests failed to generate significant relationships between both serotype and ST. In addition, we investigated associations between serotype-ST combinations and pathotype by chi-square and Fisher's exact tests and did not identify any significant associations. One possible explanation for this is the lack of discrimination due to the limitations of sample size within each subtype.

Traditional chi-square and Fisher's exact tests work best on nonsparse data (few zero values) (Mielke and Berry, 2002; Sourial et al., 2010). These tests have been used to identify associations between *S. suis* subtypes and characteristics of pathogenicity. However, most studies involved a limited number of subtypes of interest, while our study focused on all serotypes and STs identified in our sample set. Due to the diversity of the *S. suis* strains in this study and the large number of subtypes evaluated, the division of our data by pathotype resulted in sparse data. Thus, sparse data limits our ability to conduct certain analyses using common approaches for *S. suis*. An OR formula was used to evaluate statistical significance of subtype with pathotype, as well as the size of the possible effect, to limit the misidentification of associations due to sample size. For this reason, proportions were used for basic identification of relationships and OR analysis was used for further discrimination of strains.

In summary, our study increases the knowledge on *S. suis* strains circulating in the United States between 2014 and 2017 by investigating serotype and ST distributions. We identified a diverse set of strains, predominantly serotypes 1/2, 3, and 7, and as ST1, ST28, and ST94. Further investigation by pathotype classification (defined in this study) identified STs that could be differentiated as pathogenic or commensal pathotypes. The predominance of serotype 1/2 strains from clinically affected pigs in our study stresses the importance of expanding studies of virulence traits to other serotypes and STs of *S. suis*. These findings can be applied to improve the prevention and control of *S. suis* by selecting strains for diagnostics and vaccine development.

2.6 Acknowledgments

We thank the bacteriology sections at UMNVDL and KSVDL for their technical assistance. In addition, we thank Lacey Marshall-Lund (University of Minnesota, St. Paul, MN) and Marta Pérez-Sancho (Universidad Complutense de Madrid, Madrid, Spain) for technical assistance and Richard Gebhart (University of Minnesota, St. Paul, MN) and Laura Bruner (Swine Vet Center) for critical discussion of the manuscript.

2.7 Funding

The research project was funded by the Rapid Agricultural Response Fund, established by the Minnesota legislature and administered by the University of Minnesota Agricultural Experiment Station. April A. Estrada's stipend was funded by the USDA National Needs Fellowship Award, 2016-38420-25288.

2.8 Figures

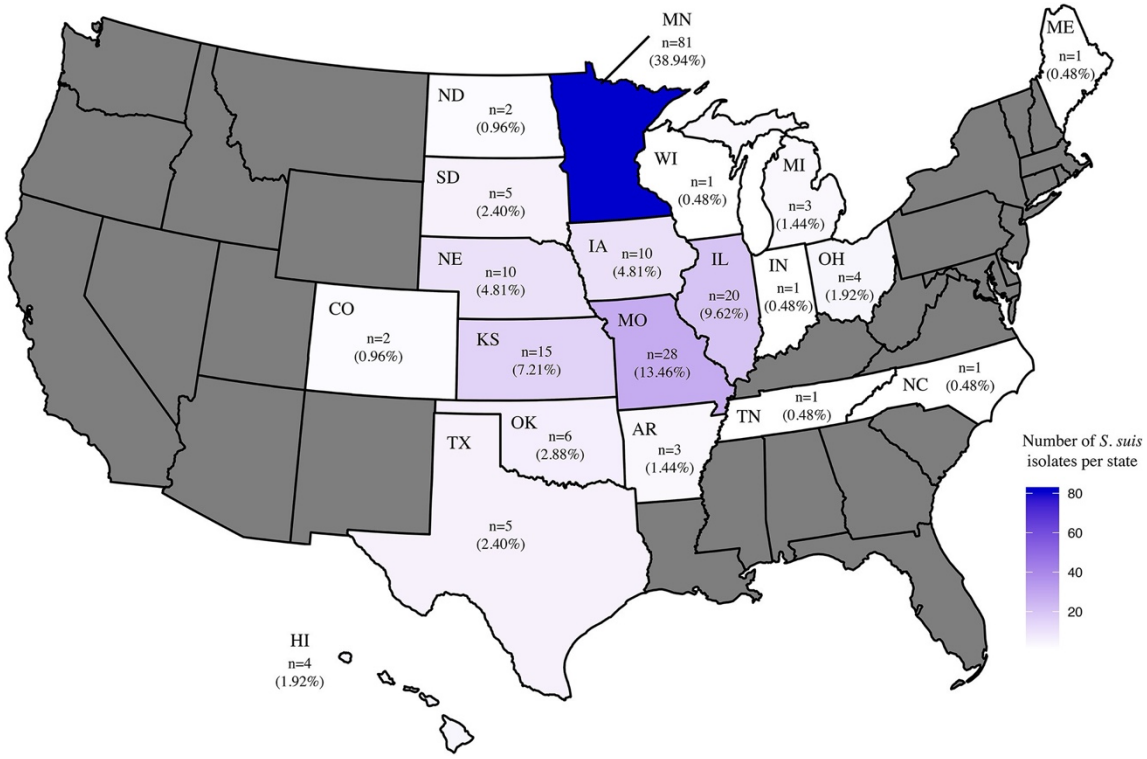

Figure 2.1 The number and percentage of *S. suis* isolates characterized from each state. States are colored according to the number of isolates characterized from each state. States without isolates are gray. Isolates from Canada (n = 4) and Mexico (n = 1) are not shown.

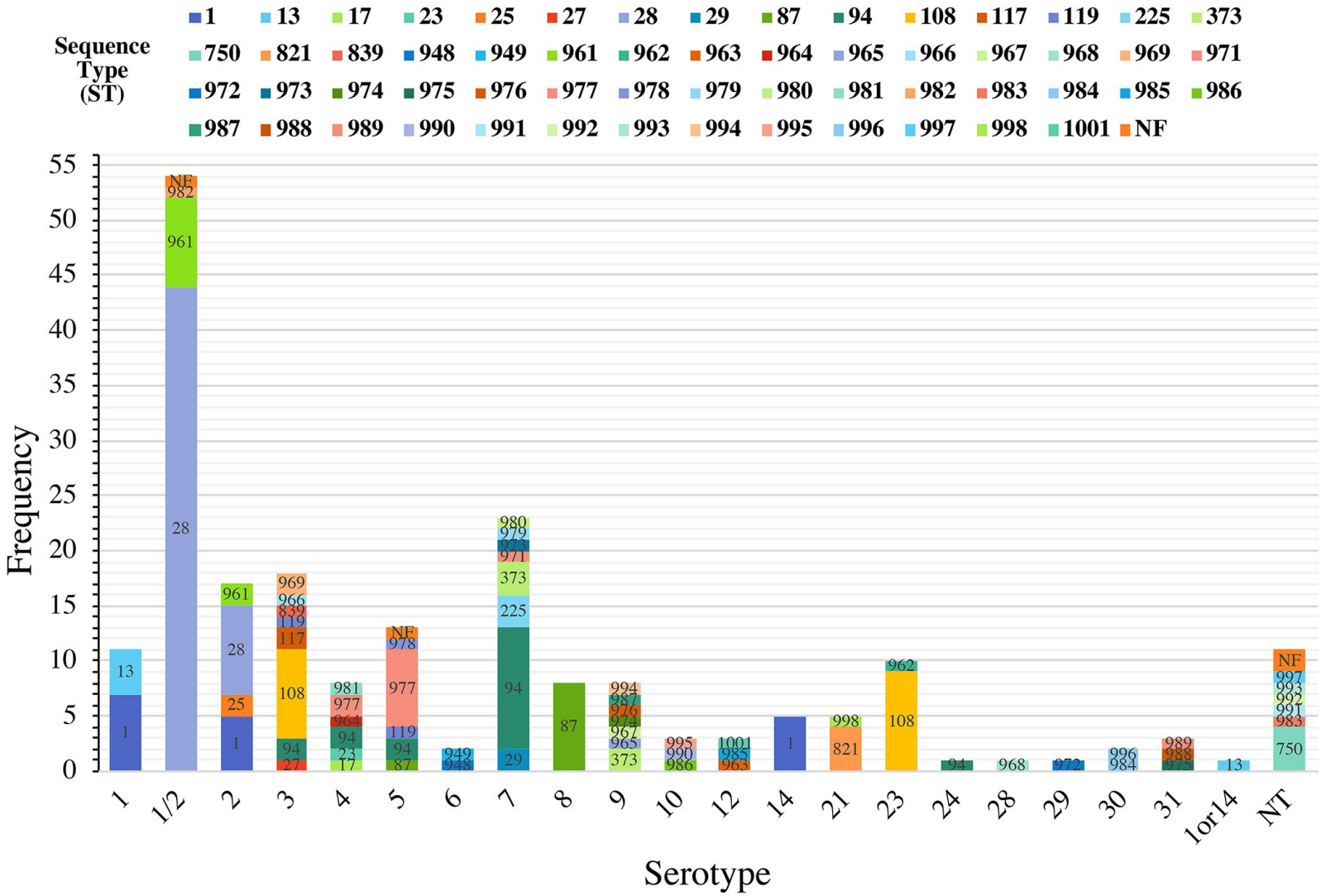
Sequence Type (ST)
1 13 17 23 25 27 28 29 87 94 108 117 119 225 373
750 821 839 948 949 961 962 963 964 965 966 967 968 969 971
972 973 974 975 976 977 978 979 980 981 982 983 984 985 986
987 988 989 990 991 992 993 994 995 996 997 998 1001 NF
Frequency
0 5 10 15 20 25 30 35 40 45 50 55
1 1/2 2 3 4 5 6 7 8 9 10 12 14 21 23 24 28 29 30 31 1or14 NT
Serotype

Figure 2.2 Distribution of *S. suis* STs by serotype. The stacked histogram illustrates the serotypes identified in this study, which were subdivided by STs. The *x* axis represents each serotype while the *y* axis represents the frequency of each serotype. Bar sections are labeled with their respective STs. The category 1or14 and (nontypeable (NT) represents isolates with serotypes that could not be differentiated by coagglutination, PCR, or WGS.

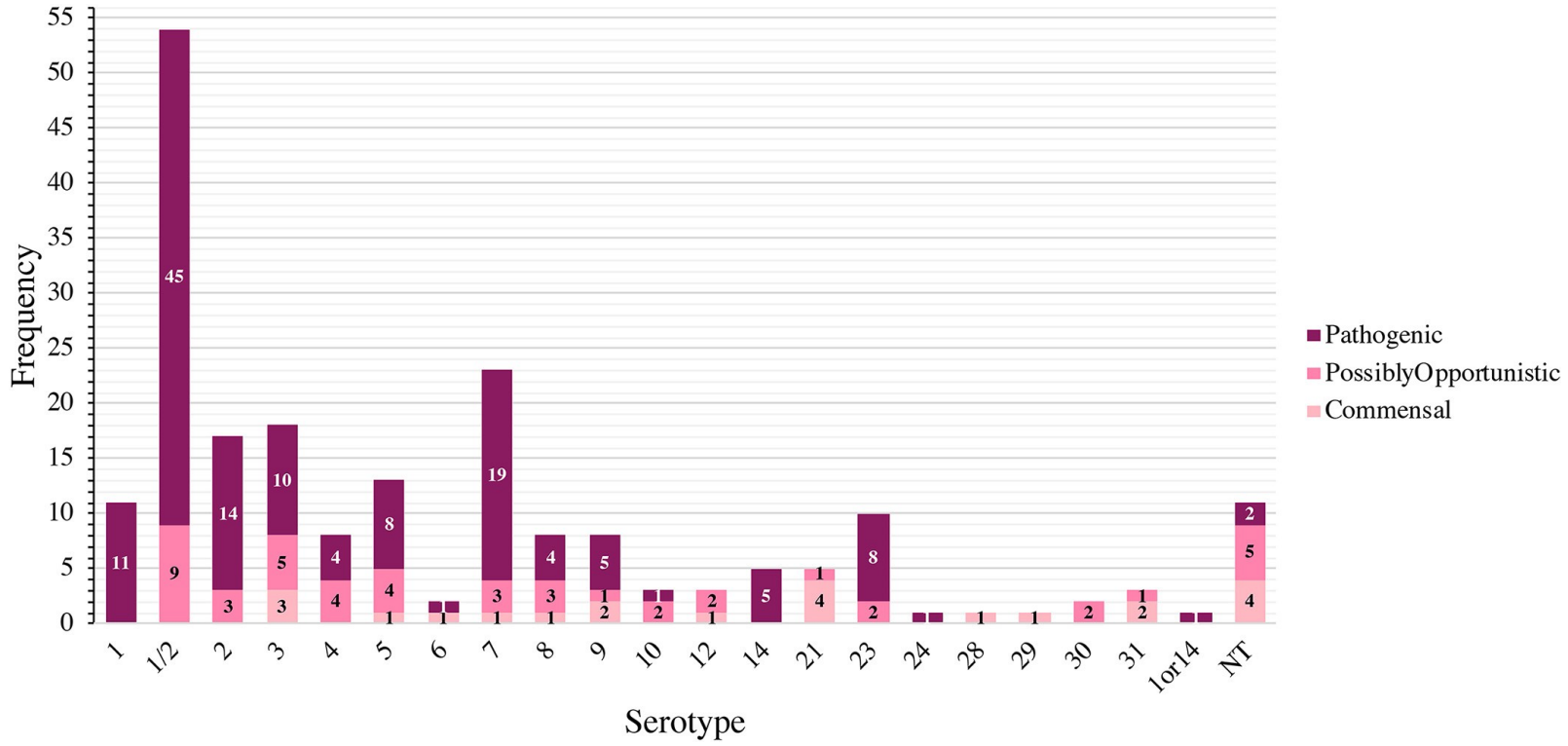

Figure 2.3 Distribution of *S. suis* pathotypes by serotype. The stacked histogram illustrates the serotypes identified in this study, which were subdivided by pathotype (pathogenic, possibly opportunistic, and commensal). The *x* axis represents each serotype while the *y* axis represents the frequency of each pathotype. Bar sections are labeled with their respective pathotypes. The category 1or14 and NT (nontypeable) represents isolates with serotypes that could not be differentiated by coagglutination, PCR, or WGS.

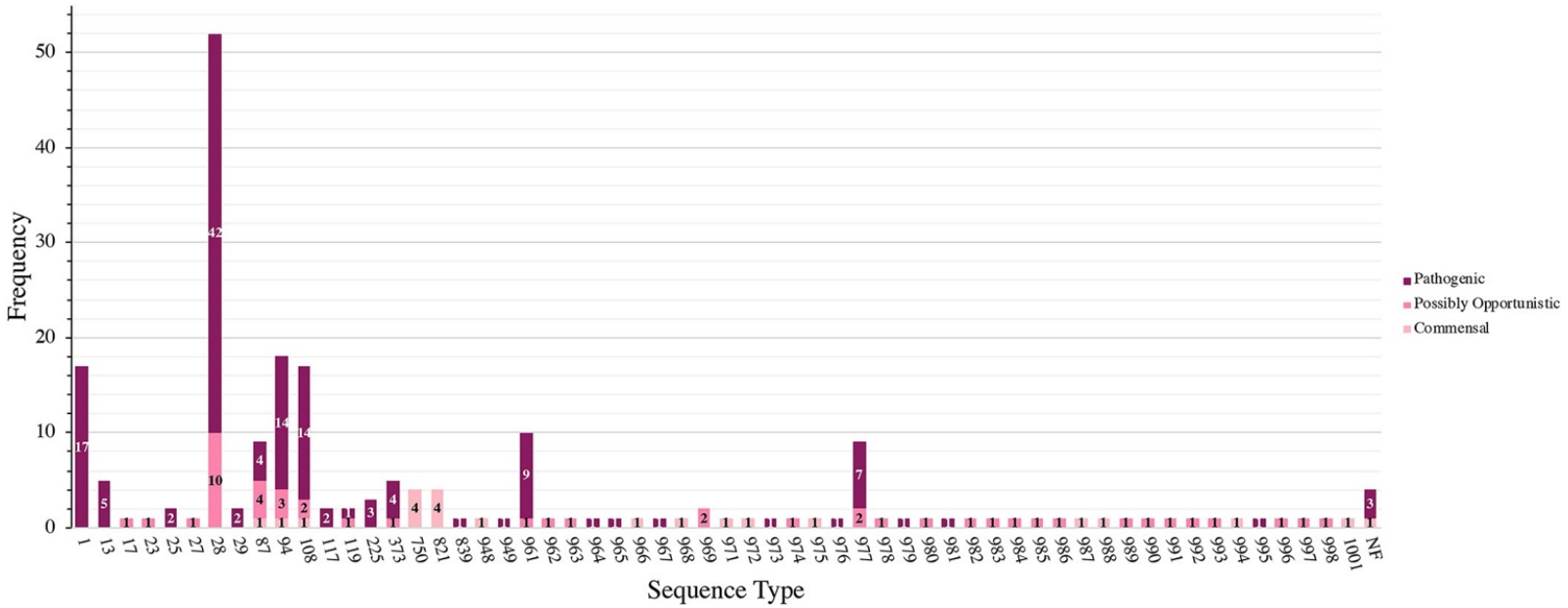

Figure 2.4 Distribution of *S. suis* pathotypes by ST. The stacked histogram illustrates the STs identified in this study, which were subdivided by pathotype (pathogenic, possibly opportunistic, and commensal). The *x* axis represents each ST while the *y* axis represents the frequency of each pathotype. Bar sections are labeled with their respective pathotypes. Not found (NF) indicates ST could not be determined because one housekeeping gene could not be identified for MLST classification.

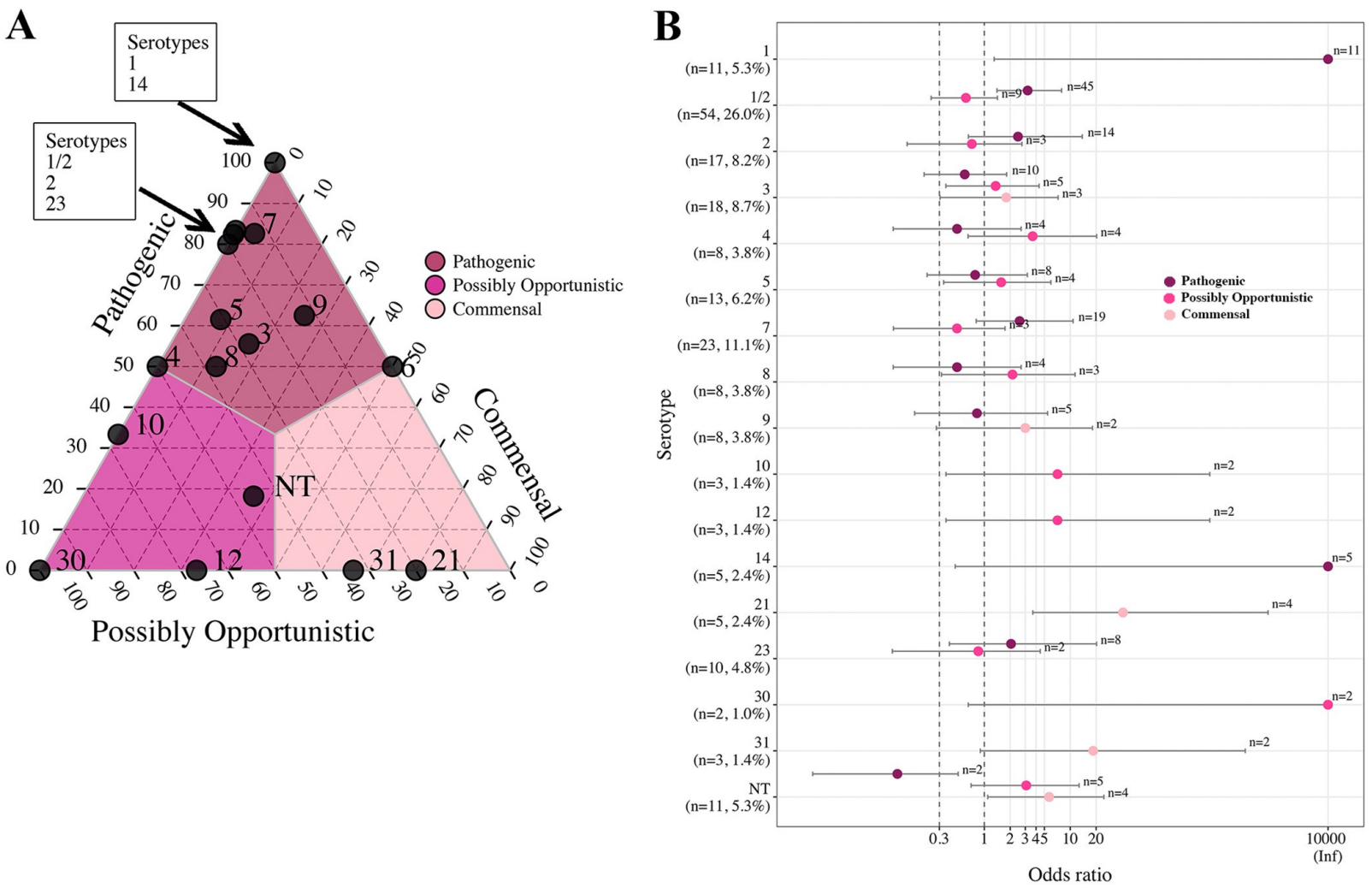
A
Serotypes
1
14
Serotypes
1/2
2
23
Pathogenic
Possibly Opportunistic
Commensal
7
5
9
3
8
4
6
10
NT
30
12
31
21
Pathogenic
Possibly Opportunistic
Commensal
B
Serotype
1 (n=11, 5.3%)
1/2 (n=54, 26.0%)
2 (n=17, 8.2%)
3 (n=18, 8.7%)
4 (n=8, 3.8%)
5 (n=13, 6.2%)
7 (n=23, 11.1%)
8 (n=8, 3.8%)
9 (n=8, 3.8%)
10 (n=3, 1.4%)
12 (n=3, 1.4%)
14 (n=5, 2.4%)
21 (n=5, 2.4%)
23 (n=10, 4.8%)
30 (n=2, 1.0%)
31 (n=3, 1.4%)
NT (n=11, 5.3%)
Pathogenic
Possibly Opportunistic
Commensal
0.3
1
2
3
4
5
10
20
10000
(Inf)
Odds ratio

Figure 2.5 Ternary and OR plots summarizing the associations between *S. suis* pathotype and serotype. Only serotypes containing more than a single isolate are illustrated in the ternary plot, and only pathotype-serotype combinations containing more than a single isolate are illustrated in the OR plot. **(A)** The proportions of isolates classified as the pathogenic, possibly opportunistic, and commensal pathotype in each of the 16 serotypes (and the NT category) were plotted. The gray lines and color shading denote pathotype boundaries. **(B)** OR plot for 16 serotypes (and the NT category) versus pathotype. The dotted lines illustrate the minimum lower limit (OR, 0.3) and typical threshold (OR, 1) for identifying significant ORs. Error bars represent the 95% confidence intervals. Inf, Infinite. Nontypeable (NT) represents isolates which could not be serotyped using coagglutination, PCR, or WGS.

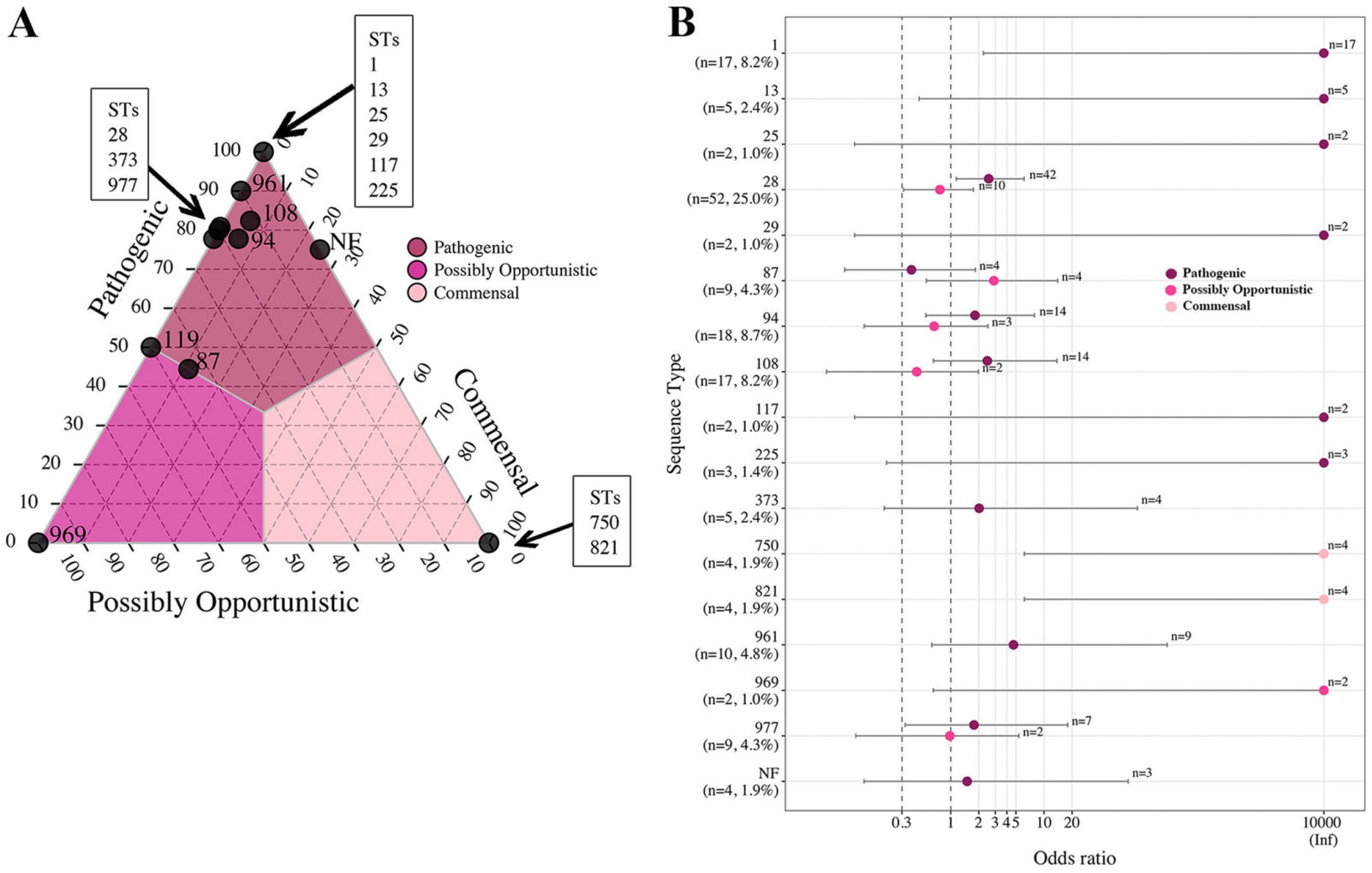
A
STs
1
13
25
29
117
225
STs
28
373
977
STs
750
821
Pathogenic
Possibly Opportunistic
Commensal
961
108
94
NF
119
87
969
B
Sequence Type
Odds ratio
Pathogenic
Possibly Opportunistic
Commensal
1 (n=17, 8.2%)
13 (n=5, 2.4%)
25 (n=2, 1.0%)
28 (n=52, 25.0%)
29 (n=2, 1.0%)
87 (n=9, 4.3%)
94 (n=18, 8.7%)
108 (n=17, 8.2%)
117 (n=2, 1.0%)
225 (n=3, 1.4%)
373 (n=5, 2.4%)
750 (n=4, 1.9%)
821 (n=4, 1.9%)
961 (n=10, 4.8%)
969 (n=2, 1.0%)
977 (n=9, 4.3%)
NF (n=4, 1.9%)
0.3
1
2
3
4
5
10
20
10000
(Inf)

Figure 2.6 Ternary and OR plots summarizing the associations between *S. suis* pathotype and ST. Only STs containing more than a single isolate are illustrated in the ternary plot, and only pathotype-ST combinations containing more than a single isolate are illustrated in the OR plot. **(A)** The proportions of isolates classified as the pathogenic, possibly opportunistic, and commensal pathotype in each of the 17 STs (and the NF category) were plotted. The gray lines and color shading denote pathotype boundaries. **(B)** OR plot for 16 STs (and the NF category) versus pathotype. ST119 contained a single isolate of each pathogenetic and possibly opportunistic pathotype failing to meet the criteria for the plot. The dotted lines illustrate the minimum lower limit (OR, 0.3) and typical threshold (OR, 1) for identifying significant ORs. Error bars represent the 95% confidence intervals. Infinite (Inf) is represented by the value 10,000 for visualization purposes. Not found (NF) indicates ST could not be determined because one housekeeping gene could not be identified for MLST classification.

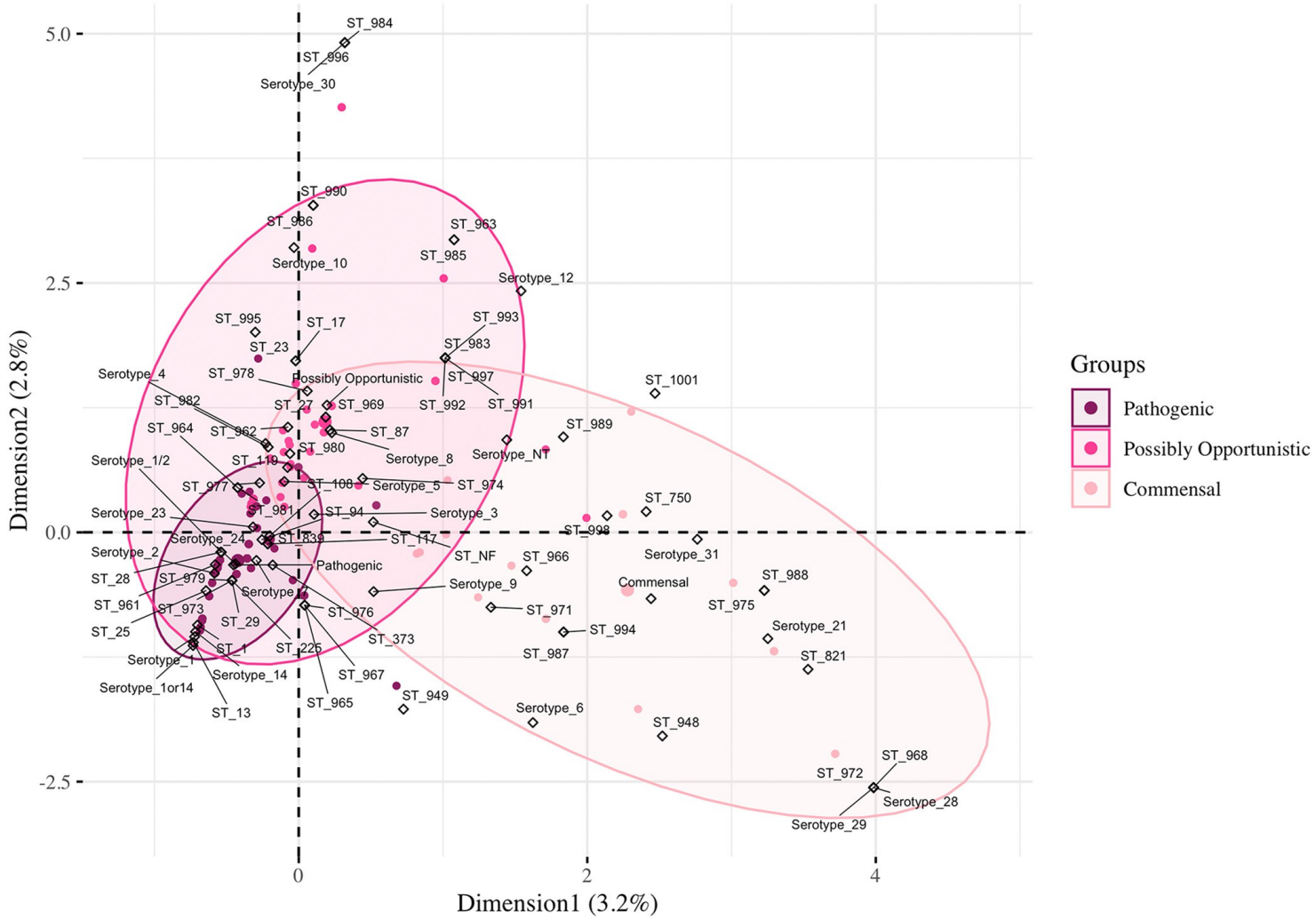
Groups
Pathogenic
Possibly Opportunistic
Commensal
Dimension1 (3.2%)
Dimension2 (2.8%)
5.0
2.5
0.0
-2.5
0
2
4
ST_984
ST_996
Serotype_30
ST_990
ST_986
Serotype_10
ST_963
ST_985
Serotype_12
ST_995
ST_17
ST_23
ST_993
ST_983
Serotype_4
ST_978
Possibly Opportunistic
ST_997
ST_1001
ST_982
ST_27
ST_969
ST_992
ST_991
ST_964
ST_962
ST_87
ST_989
ST_980
Serotype_8
Serotype_NT
Serotype_1/2
ST_119
ST_977
ST_108
Serotype_5
ST_974
ST_750
ST_981
ST_94
Serotype_3
Serotype_23
ST_998
Serotype_24
ST_838
ST_117
Serotype_31
Serotype_2
ST_NF
ST_966
ST_28
ST_979
Pathogenic
Serotype_9
Commensal
ST_988
ST_961
ST_973
Serotype_7
ST_971
ST_975
ST_25
ST_29
ST_976
ST_994
Serotype_21
ST_373
ST_1
ST_225
ST_987
ST_821
Serotype_1
Serotype_14
ST_967
Serotype_1or14
ST_949
ST_965
Serotype_6
ST_948
ST_13
ST_968
ST_972
Serotype_28
Serotype_29

Figure 2.7 Three-way MCA analyzing the associations among pathotypes, serotypes, and STs. The ellipses represent 95% of isolates in each pathotype. Dots represent isolates colored by their respective pathotype, and hollow diamonds represent the three active variables (pathotype, serotype, and ST). Nontypeable (NT) represents isolates which could not be serotyped using coagglutination, PCR, or WGS. Not found (NF) indicates ST could not be determined because one housekeeping gene could not be identified for MLST classification.

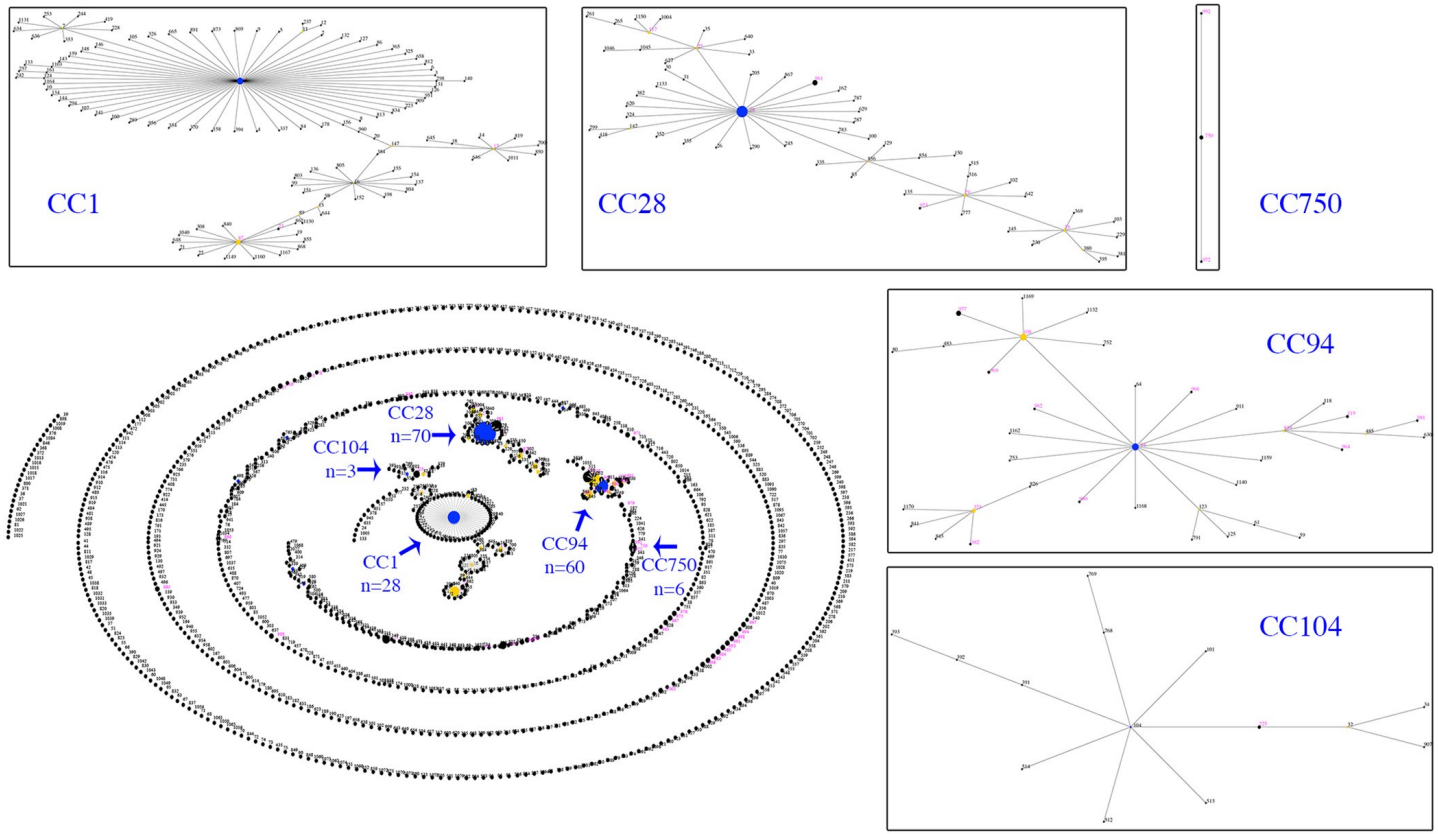

Figure 2.8 eBURST illustration of the global *S. suis* population. Primary founders (blue) are positioned at the center of the cluster and subgroup founders are shown in yellow. Clusters of linked STs correspond to CCs. Magenta dots mark the STs identified in our study and arrows mark the CCs relevant to this study. Individual CCs have been expanded to illustrate relationships among STs.

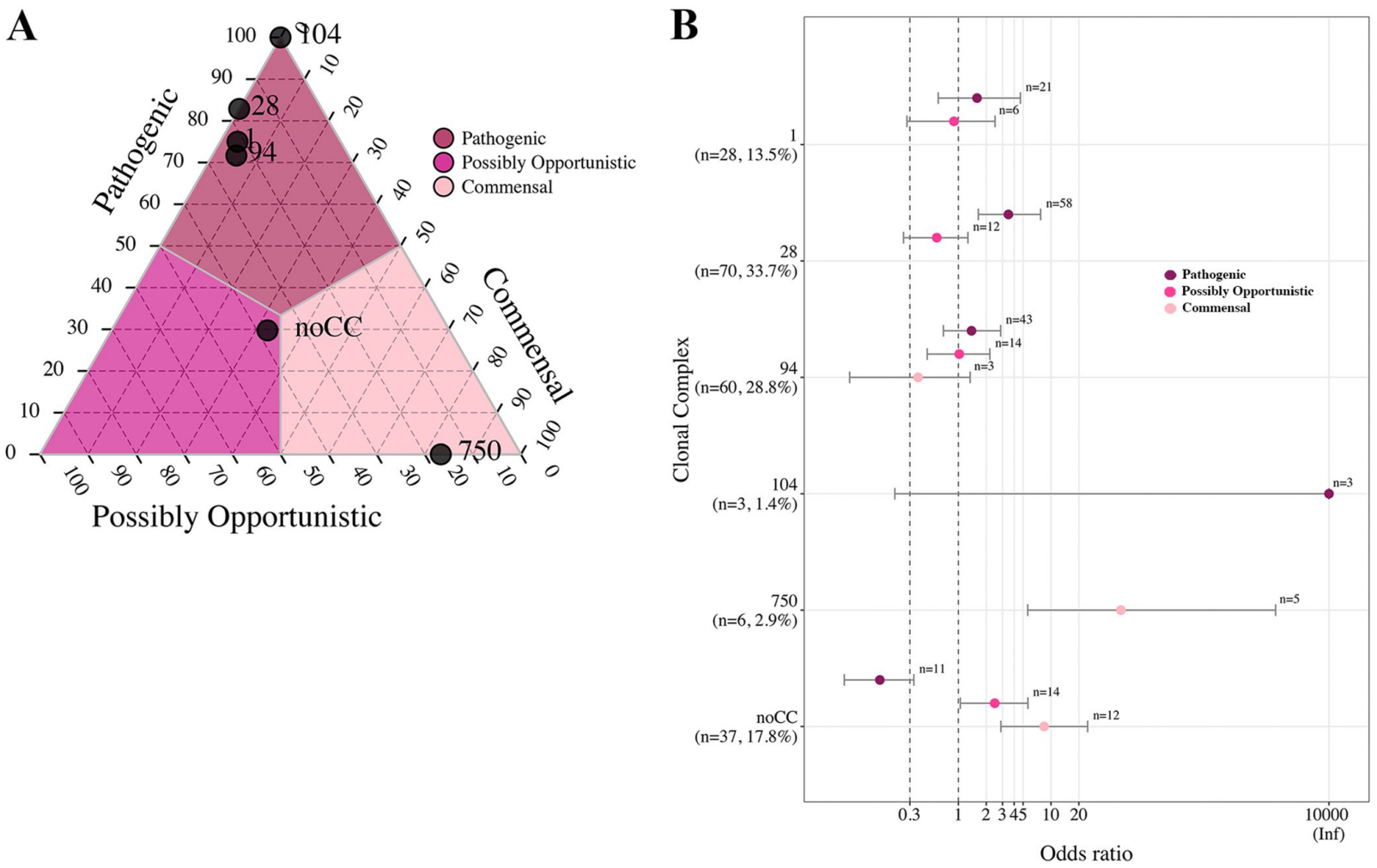
A
Pathogenic
Possibly Opportunistic
Commensal
104
28
1
94
noCC
750
B
Clonal Complex
1
(n=28, 13.5%)
28
(n=70, 33.7%)
94
(n=60, 28.8%)
104
(n=3, 1.4%)
750
(n=6, 2.9%)
noCC
(n=37, 17.8%)
n=21
n=6
n=58
n=12
n=43
n=14
n=3
n=3
n=5
n=11
n=14
n=12
Pathogenic
Possibly Opportunistic
Commensal
0.3
1
2
3
4
5
10
20
10000
(Inf)
Odds ratio

Figure 2.9 Ternary and OR plots summarizing the associations between pathotype and CC. (A) The proportions of isolates classified as the pathogenic, possibly opportunistic, and commensal pathotype in each of the five CCs (and the noCC category) were plotted. The gray lines and color shading denote pathotype boundaries. **(B)** OR plot for all five CCs (and the noCC category) versus pathotype. The dotted lines illustrate the minimum lower limit (OR, 0.3) and typical threshold (OR, 1) for identifying significant ORs. Error bars represent the 95% confidence intervals. Infinite (Inf) is represented by the value 10,000 for visualization purposes. noCC represents the group of isolates lacking a CC.

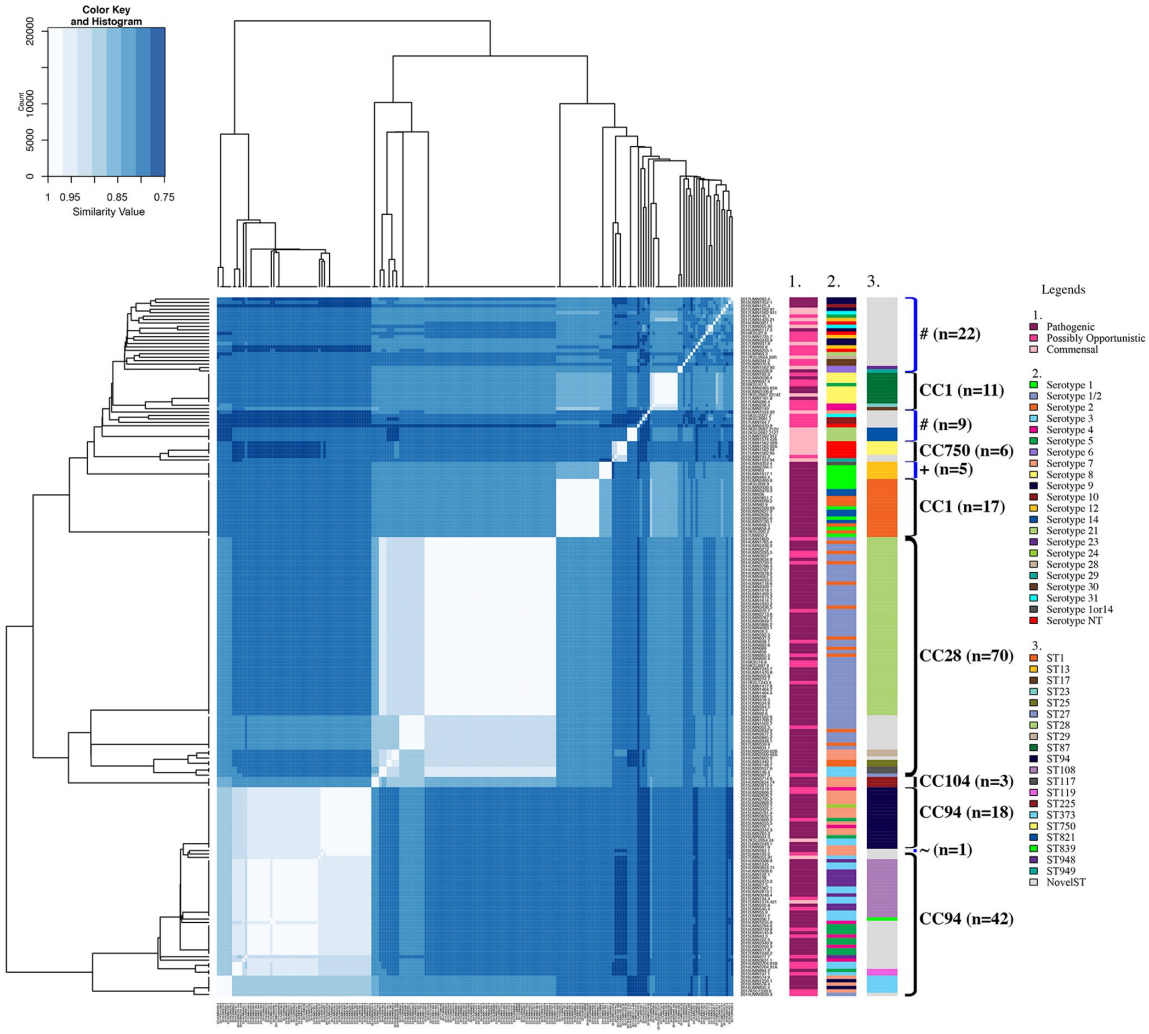

Figure 2.10 MLST sequence identity heatmap of *S. suis*. Isolates are annotated (colored rectangles) by pathotype (1), serotype (2), and ST (3). The five CCs are indicated by black brackets, with the number of isolates in the CC. Blue brackets represent clusters of isolates without a CC. Nontypeable (NT) represents isolates which could not be serotyped using coagglutination, PCR, or WGS. #, group of isolates lacking a CC; +, ST13 not within a CC but closest to CC1; ~, ST979 not within a CC but closest to CC94.

2.9 Tables

Table 2.1 Distribution of *S. suis* isolates by pathotype classification and tissue of origin

Pathotype	Tissue of origin	*n*
Pathogenic ($n = 139$, 66.8%)	Brain/meninges/spinal cord	56
	Heart	25
	Joint/synovial fluid	23
	Liver	35
Possibly opportunistic ($n = 47$, 22.6%)	Lung	45
	Nasal	2
Commensal ($n = 22$, 10.6%)	Laryngeal	1
	Nasal	17
	Tonsil	4

Table 2.2 Primary *S. suis* CCs identified in this study

No. of CCs	Total no. of isolates[a]	No. of pathogenic isolates	No. of possibly opportunistic isolates	No. of commensal isolates	ST(s)
1	28	21	6	1	1, 17, 23, 87
28	70	58	12	0	25, 27–29, 117, 961[b], 973[b]
94	60	43	14	3	94, 108, 119, 373, 839, 962[b], 964[b], 966[b], 969[b], 977[b], 980–982[b]
104	3	3	0	0	225
750	6	0	1	5	750, 972[b], 992[b]
noCC[c]	37	11	14	12	13, 821, 948, 949, 963[b], 965[b], 967[b], 968[b], 971[b], 974–976[b], 978[b], 979[b], 983–991[b], 993–998[b], 1001[b]

[a]STs could not be determined for 4 of the 208 isolates because 1 housekeeping gene could not be identified for MLST classification.

[b]Novel ST(s).

[c]CCs were lacking for STs that occurred as singletons or had no determined founder.

CHAPTER 3

Proposed virulence-associated genes of *Streptococcus suis* isolates from the United States serve as predictors of pathogenicity

April A. Estrada [1]*, Marcelo Gottschalk[2], Aaron Rendahl[1], Stephanie Rossow[3], Lacey Marshall-Lund[3], Douglas G. Marthaler[4] and Connie J. Gebhart[1]

[1]Department of Veterinary and Biomedical Sciences, College of Veterinary Medicine, University of Minnesota, Saint Paul, MN, USA

[2]Faculty of Veterinary Medicine, University of Montreal, Saint-Hyacinthe, Quebec, Canada

[3]Veterinary Diagnostic Laboratory, College of Veterinary Medicine, University of Minnesota, Saint Paul, MN, USA

[4]Veterinary Diagnostic Laboratory, College of Veterinary Medicine, Kansas State University, Manhattan, KS, USA

Estrada AA et al. Proposed virulence-associated genes of *Streptococcus suis* isolates from the United States serve as predictors of pathogenicity. *Porc Health Manag*. 2021;7(22). doi:10.1186/s40813-021-00201-6

3.1 Synopsis

There is limited information on the distribution of virulence-associated genes (VAGs) in U.S. *Streptococcus suis* isolates, resulting in little understanding of the pathogenic potential of these isolates. This lack also reduces our understanding of the epidemiology associated with *S. suis* in the United States and thus affects the efficiency of control and prevention strategies. In this study we applied whole genome sequencing (WGS)-based approaches for the characterization of *S. suis* and identification of VAGs. Of 208 *S. suis* isolates classified as pathogenic, possibly opportunistic, and commensal pathotypes, the genotype based on the classical VAGs (*epf*, *mrp*, and *sly* encoding the extracellular protein factor, muramidase-release protein, and suilysin, respectively) was identified in 9% (*epf*+/*mrp*+/*sly*+) of the pathogenic pathotype. Using the chi-square test and LASSO regression model, the VAGs *ofs* (encoding the serum opacity factor) and *srtF* (encoding sortase F) were selected out of 71 published VAGs as having a significant association with pathotype, and both genes were found in 95% of the pathogenic pathotype. The *ofs*+/*srtF*+ genotype was also present in 74% of 'pathogenic' isolates from a separate validation set of isolates. Pan-genome clustering resulted in the differentiation of a group of isolates from five swine production companies into clusters corresponding to clonal complex (CC) and virulence-associated (VA) genotypes. The same CC-VA genotype patterns were identified in multiple production companies, suggesting a lack of association between production company, CC, or VA genotype. The proposed *ofs* and *srtF* genes were stronger predictors for differentiating pathogenic and commensal *S. suis* isolates compared to the classical VAGs in two sets of U.S. isolates. Pan-genome analysis in combination with metadata (serotype, ST/CC, VA genotype) was illustrated to be a valuable subtyping tool to describe the genetic diversity of *S. suis*.

Keywords *Streptococcus suis*, Virulence-associated genes (VAGs), Pathotype, Pan-genome, Genetic diversity

3.2 Introduction

The severe clinical presentation associated with *Streptococcus suis* infection is of increasing concern in the U.S. swine industry. The heterogeneity of *S. suis* can be described by serotyping and multi-locus sequence typing (MLST), and currently 29 true serotypes (1–19, 21, 23–25, 27–31, and 1/2) and 1551 registered sequence type (ST) profiles (as of December 2020) exist (Higgins et al., 1995; King et al., 2002; Hill et al., 2005; Tien et al., 2013; Jolley et al., 2018). The numerous *S. suis* serotypes and STs limit our attempts to understand the epidemiology of the disease in an effort to prevent and manage the various clinical manifestations. Further*, S. suis* has zoonotic potential, and many of the effective antibiotics available for treatment are of high or critically important status per U.S. Food and Drug Administration's Guidance for Industry #152 (2003). Also, serotype variations make it difficult to compare isolates within and across geographically distinct pig populations. The development of effective universal vaccines is hindered by the number of different virulent serotypes and the lack of knowledge of serotype- or ST-specific virulence markers and associated clinical disease.

Historically, systematic characterization of *S. suis* isolates occurred more extensively in other countries compared to the United States. For instance, serotypes 2, 3, and 1/2 have been well-characterized in Canadian swine populations (Higgins and Gottschalk, 2001; Messier et al., 2008; Gottschalk et al., 2013). In addition, virulence-associated genes (VAGs) (*epf*, *mrp*, and *sly*) and STs indicative of virulence potential were identified, mostly for serotype 2 (Gottschalk et al., 2013, 1998; Fittipaldi et al., 2011, 2009). Experimental studies illustrating the virulence potential of Canadian serotype 2 strains (ST1, ST25, and ST28) suggest the virulence potential of ST28 strains is low at best in Canada, but a very different clinical presentation was being observed on U.S. swine farms with ST28 (Fittipaldi et al., 2011; Athey et al., 2015; Auger et al., 2016a). In the past 4 to 5 years, *S. suis* infections on U.S. swine farms appeared to be more persistent and severe (Leuwerke et al., 2019; Pig Health Today, 2019). However, whether this is a result of new circulating strains, an increase in virulence, or some other

cause is not well understood, reinforcing the importance of the characterization of *S. suis* to monitor changes in strains within a herd.

U.S. swine practitioners utilize herd vaccination strategies as a means of controlling *S. suis* disease. However, selecting representative isolates and properly timing the administration of vaccines still remain a challenge (Baums et al., 2010). As a result of the diversity of *S. suis*, limited commercial vaccines are available, and many practitioners develop and maintain farm-specific autogenous vaccines. In addition, clear criteria for identifying pathogenic strains that cause primary disease are lacking, making isolate selection for vaccines more difficult (Segura et al., 2017). Isolates are commonly selected based on criteria such as serotype and isolation from systemic tissues (Galina et al., 1992; Reams et al., 1994). However, due to the diversity within and between serotypes, cross-protection between, and even within, different serotypes is difficult to attain (de Greeff et al., 2011; Zhang et al., 2011; Quessy et al., 1994; Segura et al., 2020). Moreover, the presence of virulence markers is critical for selecting isolates for autogenous vaccines. Over 100 putative and confirmed virulence factors and markers (not crucial or critical for virulence) for *S. suis* have been described in the literature, but few have been verified in experimental models (Fittipaldi et al., 2012; Segura et al., 2017). These include Eurasian serotype 2 virulence markers extracellular protein factor (*epf* gene), muramidase-released protein (*mrp* gene), and suilysin (*sly* gene), which have been investigated in STs 1, 25, and 28 strains from North America (Gottschalk and Segura, 2000; Fittipaldi et al., 2011; Goyette-Desjardins et al., 2014).

The application of genomic approaches to identify associations between VAGs and disease manifestation can lead to a better understanding of *S. suis* pathogenesis. However, a comparative genomic study investigating the current distribution of *S. suis* VAGs in U.S. isolates is lacking. Recently, we reported associations of various pathotypes with subtypes, including serotype and ST, of *S. suis* (Estrada et al., 2019). In this current study, a genomic approach was utilized to identify associations between VAGs and pathotype of U.S. isolates while evaluating the likelihood of classical Eurasian serotype 2 VAGs and newly proposed VAGs to identify pathogenic strains. In addition,

pan-genome genetic relationships, along with their VAG profiles (virulence-associated genotypes), were investigated for isolates within and between swine production companies. Finally, we applied the genomic approach for identifying associations between VAGs and pathogenicity classification to a validation set of *S. suis* isolates to determine whether our approach was robust enough to identify pathogenic strains isolated from other swine production companies.

3.3 Materials and methods

3.3.1 Source of isolates and collection of epidemiological data

A training set of 208 *S. suis* isolates were used in this study. These isolates, previously described by Estrada et al. (2019), were classified into three pathotypes (pathogenic, possibly opportunistic, and commensal) based on clinical information and site of isolation. "Pathogenic" isolates were obtained from systemic tissues such as the brain/meninges and heart. "Possibly opportunistic" isolates were predominantly from lung samples from pigs without signs of neurological or systemic disease. "Commensal" isolates were from laryngeal, tonsil, or nasal samples retrieved from farms with no current control methods for *S. suis* disease.

Furthermore, epidemiological data, such as swine production company and site, were collected for the training set of isolates. The swine production companies coded as A, D, E, K, and M are all large operations that range in size from 70,000 to 340,000 sows and with headquarters in the United States (A and D = MN, E = MO, K = KS, and M = IL).

3.3.2 VAG profiling

VAG profiling was performed on the training set using a custom database of previously published VAGs of *S. suis* (Table S3.1) (Estrada et al., 2019). Illumina sequencing reads were mapped to reference DNA sequences (≥ 60% coverage and ≥ 90% sequence identity) using the SRST2 (Short Read Sequence Typing for Bacterial Pathogens) program (Inouye et al., 2014). The construction of a presence and absence

heatmap (Euclidian distances and UPGMA clustering) was performed with R software (R Core Team, 2017).

3.3.3 Statistical analysis

Associations between published *S. suis* VAGs and pathotype, as previously defined by Estrada et al. (Estrada et al., 2019), were investigated. Published VAGs present in a majority of isolates (> 90%, 188/208) were removed. VAGs present in < 50% of the pathogenic pathotype (< 70/139) were also removed. Remaining VAGs were tested by chi-square, comparing the three pathotypes and the status (presence/absence) of individual genes. Genes lacking a significant (chi-square p-value < 0.05) association with pathotype were removed from the analysis. The remaining genes were analyzed using the Least Absolute Shrinkage and Selection Operator (LASSO) regression model (Tibshirani, 1996).

The LASSO regression model reduces coefficients to zero and gradually eliminates genes that have no or low correlation with the target variable. The LASSO model was used to determine VAGs that may serve as the 'best' predictors of pathogenicity, in this case using the pathogenic pathotype as the indicator of pathogenicity. The analysis was performed using the R package glmnet and the best lambda penalty value to determine the fewest number of predictor genes (Friedman et al., 2010). Due to variation in the number of predictor VAGs in each run, we ran 100 iterations of the LASSO model to determine the most relevant predictor genes. Predictor VAGs reported in all 100 iterations were considered relevant candidate VAGs.

3.3.4 Genome assembly and pan-genome analysis

Genome assembly was performed on Illumina sequencing data from the training set (Estrada et al., 2019): SRA accession numbers SRR9123061-SRR9123268. Genome assemblies were generated using MEGAHIT *de-novo* assembler (k-mer range of 25–225) and polished using Pilon (Walker et al., 2014; Li et al., 2016). QUAST was used to evaluate the genome assemblies (Gurevich et al., 2013). Only contigs that were 500 bp or larger were kept for annotation by Prokka to predict coding sequences (Seemann, 2014).

The pan-genome was annotated using Roary with a 90% BLASTp identity cut-off to define clusters of genes and allowing paralog clustering (Page et al., 2015; van Vliet, 2017). The FastTree program was used to generate an approximately-maximum-likelihood phylogenetic tree based on the binary presence and absence of core and accessory genes. Percent similarity was calculated as the percentage of shared genes in the pan-genome.

3.3.5 Selection and whole genome sequencing of validation set

Thirty-two isolates obtained from a single swine production company from 2017 to 2019 were classified as either 'pathogenic' or of 'unknown-pathogenicity' based on tissue source (Table S3.2). Isolates classified as 'pathogenic' were obtained from the brain (n = 19). The isolates of 'unknown-pathogenicity' were isolated from non-systemic tissues (no neurological signs) (n = 13). The *S. suis* isolates were sequenced and the sequencing reads were processed using a similar method as described for the training set (Estrada et al., 2019). Isolates were confirmed as *S. suis* if they possessed the *S. suis*-specific recombination/repair protein (*recN*) sequence (*Streptococcus suis* 05HAS68, Accession CP002007).

3.3.6 Serotype, MLST, VAG profile, and pan-genome analysis of validation set

The serotyping of the validation set of *S. suis* isolates was verified using a *S. suis* serotyping pipeline described by Athey et al. (2016) to differentiate serotypes 2 and 1/ 2 and serotypes 1 and 14 (Athey et al., 2016). *In-silico* MLST analysis was performed using the SRST2 program, and the ST allele sequences and profiles obtained from the *S. suis* MLST database (Jolley et al., 2018). Novel STs were further grouped into major clonal complexes (CCs) as previously described (Estrada et al., 2019). Data on presence or absence of the classical VAGs (*epf*, *mrp,* and *sly*) was obtained for each of the 32 isolates as described above for the training set. Similar genome assembly and pan-genome analysis as described for the training set were performed on the 32 isolates. The number of gene clusters identified for the training and validation sets may differ due to gene duplication, pseudogenes, gene acquisition/loss, and other genomic variations, as

well as differences in the number of genomes included in the pan-genome analysis (Medini et al., 2005; Meng et al., 2017).

3.3.7 Data availability

The datasets generated and/or analyzed during the current study are available in the Sequence Read Archive. Illumina sequencing data from the training set (Estrada et al., 2019) were deposited under the accession numbers SRR9123061-SRR9123268. Illumina sequencing data from the validation set were deposited under the accession numbers SRR12964194-SRR12964227. The accession numbers from the validation set can also be found in Table S3.2.

3.4 Results

3.4.1 VAG profiling

Distribution of the epf, mrp, and sly genes

In our previous study of 208 *S. suis* isolates (referred to as the training set), 139, 47, and 22 were classified as the pathogenic, possibly opportunistic, and commensal pathotype, respectively (Estrada et al., 2019). The training set was also characterized by determination of serotype, MLST, and CC. In the current study, the distribution of the *epf*, *mrp*, and *sly* genes was bioinformatically determined for the 208 isolates in the training set. These classical VAGs *epf*, *mrp*, and *sly* were identified in 20 (14.4%), 127 (91.4%), and 77 (55.4%) isolates of the pathogenic pathotype, respectively (Table 3.1). The *epf* gene was predominantly present in serotypes 1, 2, and 14 and CC1 isolates while *mrp* and *sly* were distributed among a diverse set of subtypes. The *epf*, *mrp*, and *sly* genes were identified in 0 (0%), 6 (27.3%), and 4 (18.2%) isolates of the commensal pathotype, respectively (Table 3.2). We further investigated genotype combinations of the *epf*, *mrp*, and *sly* genes and their distributions in STs 1, 25, and 28 (Table 3.2). The predominant genotype in the pathogenic pathotype was *epf*−/*mrp*+/*sly*- (41.0%, 57/139) followed by *epf*−/*mrp*+/*sly*+ (36.0%, 50/139). A majority of the ST28 (94%) and both ST25 isolates in the training set possessed the *epf*−/*mrp*+/*sly*- genotype. The *epf*+/*mrp*+/ *sly*+ genotype

was identified in only 20 of the 139 isolates classified as the pathogenic pathotype. All 17 ST1 isolates possessed the *epf*+/*mrp*+/*sly*+ genotype. In summary, a majority of isolates, even those of the pathogenic pathotype, lacked the three classical VAGs, but all the isolates containing the three VAGs were classified as pathogenic or ST1.

Determining predictors of pathotype by VAG profiling

Given the limited distribution of classical VAGs (*epf* and *sly*) among isolates in the pathogenic pathotype, the classical VAGs are not appropriate indicators of pathogenicity for non-serotype 2 *S. suis* isolates from the United States. Thus, a total of 71 previously published *S. suis* VAGs (including *epf*, *mrp*, and *sly*) were investigated for the presence of alternative genes that may be indicators of pathogenic strains. Thirty-two (45%) VAGs were present in all genomes regardless of pathotype and were clearly not indicators of the pathogenic pathotype (Table 3.3 & Table S3.3). Five VAGs were absent in all of the isolates in the commensal pathotype. *SalK* and *salR*, which encode the SalK/SalR two-component signal transduction system (Li et al., 2008), were not detected in any of the isolates despite mapping to different reference sequences.

Clustering analysis was used to determine if relationships between the presence of previously published VAGs and pathotype existed. The analysis of 71 VAGs identified three clusters (Cluster I-III), two of which associated with pathotype (Fig. 3.1 & Table S3.4). Cluster I consisted of isolates of all three pathotypes. Cluster II predominantly consisted of isolates from the pathogenic pathotype and lacked isolates from the commensal pathotype. Cluster III contained the majority of isolates from the commensal pathotype (73%). Isolates in the pathogenic cluster (Cluster II) were predominantly characterized as serotype 1/2 CC28. Serotype 1/14 CC1 isolates formed a subcluster of Cluster I which lacked isolates from the commensal pathotype. Clustering analysis also illustrated multiple candidate published VAGs for discriminating between pathotypes, specifically VAGs present in the two pathogenic clusters and absent in the commensal cluster.

We then performed statistical analyses to test for associations between VAGs and pathotype. Of the 71 published VAGs detected in the genomes, 16 were tested by chi-square and 14 were considered significant (chi-square $p < 0.05$) (Table 3.3 & Table S3.5). The classical VAGs *mrp* and *sly* were considered significant by chi-square. The 14 VAGs that were significant by chi-square were further analyzed by the LASSO model. The *sly* gene was in the top ten VAGs identified by LASSO, but *mrp* was not [data not shown]. The LASSO model identified four other candidate VAGs associated with the pathogenic pathotype (Table 3.4). The VAGs *ofs* and *srtF* were present in over 95% (≥ 132/139) of isolates in the pathogenic pathotype and thus, the presence of both genes was tested as predictors of the pathogenic pathotype. Ninety-five percent (132/139) of the pathogenic pathotype contained both genes while only 23% (5/22) of the commensal pathotype contained both genes.

3.4.2 Diversity of U.S. *S. suis* by pan-genome analysis

Relatedness of 208 S. suis isolates by pan-genome analysis

Pan-genome analysis of 208 *S. suis* genomes generated a pan-genome of 8373 gene clusters and illustrated multiple clusters that corresponded to the five MLST CCs (CC1, CC28, CC94, CC104, and CC750) (Fig. 3.2) (Estrada et al., 2019). Isolates from at least sixteen swine production companies (A-P) (≥ 2 isolates each) were identified in the data set, with A (n = 13), D (n = 16), E (n = 18), K (n = 21), and M (n = 16) representing the five production companies with the most isolates in this study (predominant production companies). The most predominant CCs (CC1, CC28, and CC94) were identified in multiple production companies. CCs 1, 28, and 94 were identified in 12, 11, and 12 of the 16 production companies, respectively.

Relatedness of isolates within the five predominant production companies

The genetic relationships between pathogenic and possibly opportunistic isolates within a production company were investigated using pan-genome analysis for each of the five predominant production companies A, D, E, K, and M (Fig. 3.3). None of the isolates from these production companies were classified as the commensal pathotype

(Estrada et al., 2019). In addition, we explored associations between pan-genome clusters and genotypes of the classical (*epf*, *mrp*, and *sly*) and proposed pathogenic (*ofs* and *srtF* genes) VAGs. The isolates demonstrated various genotypes of classical VAGs, and a majority (96.4%) possessed the proposed *ofs*+/*srtF*+ genotype for predicting pathogenic strains.

Due to the diversity of isolates within each production company, we investigated pan-genome clusters and genotypes for each predominant production company. The isolates originating from production companies A, E, and M were placed into two, one, and three clusters, respectively, and multiple singletons each and demonstrated an overall 83.1–99.9%, 89.3–100% and 85.3– 99.9% similarity, respectively. A majority of the isolates originating from A and E (A = 54%, E = 72%), and many from M (50%) possessed the classical VAG *mrp* but lacked the *epf* and *sly* genes. The isolates originating from production companies D and K were placed into three clusters and multiple singletons each and demonstrated an overall 89.4–100% and 86.6.-99.9% similarity, respectively. Multiple isolates from D and K (D = 44%, K = 48%) possessed *mrp* and *sly* but lacked the *epf* gene. A few isolates from each production company (n =1 from companies D, E, K, and M, n = 2 from company A) possessed all three classical VAGs *epf*, *mrp*, and *sly*. A majority of isolates in all five production companies (A = 92.3%, D = 100%, E = 100%, K = 95.2%, M = 93.8%) possessed the proposed *ofs+/srtF+* genotype, indicating the presence of both *ofs* and *srtF* genes are a better predictor of pathogenicity than the presence of the *epf, mrp,* and *sly* genes.

Relatedness of commensal isolates

We further investigated the genetic relationships between the 22 isolates of the commensal pathotype. An 82.6–99.9% similarity was observed, with isolates forming two large clusters and multiple sub-clusters (Fig. 3.4). Thirteen isolates lacked a CC, while one, three, and five isolates were assigned to CC1, CC94, and CC750, respectively. The CC1 and CC94 isolates possessed more VAGs than the other commensal isolates with all possessing *mrp* and *sly*, and both *ofs* and *srtF,* while a majority of commensal isolates (77.3%) lacked the classical and proposed pathogenic VAGs.

Characterization of the validation set of S. suis isolates

A distinct validation set of 32 *S. suis* isolates was obtained from a single production company to perform pan-genome analysis and further test the novel proposed pathogenic genotype (*ofs+/srtF+*). These isolates were classified as either 'pathogenic' or of 'unknown-pathogenicity.' The pan-genome consisted of 7078 gene clusters among the 32 genomes, and these pan-genome clusters associated with the 'pathogenic' and 'unknown-pathogenicity' classifications, as well as with virulence-associated genotypes (Fig. 3.5). Clusters c-f corresponded to the 'pathogenic' classification. Only the isolates in clusters e and f possessed the classical VAGs *epf*, *mrp*, and *sly*. Moreover, all the isolates in these two clusters possessed the proposed pathogenic *ofs+/srtF+* genotype. A majority of the isolates in cluster d (67%) possessed the *ofs+/srtF+* genotype.

Cluster a and singletons g-n corresponded to the 'unknown-pathogenicity' classification (Fig. 3.5). A majority (86%) of the isolates in cluster a possessed the classical VAGs *mrp* and *sly*, but three isolates (43%) also possessed the proposed pathogenic genotype. A majority (75%) of the singletons g-n lacked both the classical and proposed VAGs. Two isolates possessed the proposed pathogenic genotype. The diversity and lack of VAGs in clusters g-n is similar to the diversity seen among the commensal pathotype, suggesting these isolates are commensal strains.

Supplementary material

Supplemental material for this chapter may be found at doi 10.1186/s40813-021-00201-6.

3.5 Discussion

In this study, 71 published *S. suis* VAGs (including the classical VAGs *epf*, *mrp*, *sly*) were evaluated to identify pathogenic isolates associated with systemic and neurological disease from the United States. Notably, VAGs *ofs* and *srtF* demonstrated stronger associations with the pathogenic pathotype than the other 69 VAGs, suggesting novel published VAGs associated with pathogenicity. A genotyping scheme consisting of these two genes (*ofs+*/*srtF+* genotype) identified pathogenic isolates in a validation set of

S. suis isolates, demonstrating its potential application for predicting pathogenicity in other swine production companies. The genetic diversity of isolates within and between swine production companies was evaluated by pan-genome analysis, and important associations were observed among pan-genome clusters, CCs, and virulence-associated (VA) genotypes.

Muramidase-released protein has been associated with enhanced survival of *S. suis* in human blood and an increase in blood-brain barrier permeability in mice while suilysin plays a role in the inflammatory response although neither of which have been described as being critical as virulence factors (Smith et al., 1996; Allen et al., 2001; Lun et al., 2003; Wang et al., 2015; Pian et al., 2016; Tenenbaum et al., 2016). The *epf* gene was identified in only 14% and the *sly* gene was identified in 55% of isolates in the pathogenic pathotype. The *mrp* gene was identified in 91% of the pathogenic pathotype and 27% of the commensal pathotype suggesting the classical VAG *mrp* continues to be an adequate identifier of pathogenic strains. The *epf*+/*mrp*+/*sly*+ genotype is correlated with *S. suis* clinical disease caused by European and Asian ST1 strains belonging to serotypes 1, 2, 9, and 14 (Wisselink et al., 2000; Wei et al., 2009; Goyette-Desjardins et al., 2014). The ST1 isolates in the training set had the *epf*+/*mrp*+/*sly*+ genotype, and the ST25 and ST28 isolates had the *epf*–/*mrp*+/*sly*- genotype, confirming the use of the classical VAGs for identifying virulent ST1 strains but the limited use for identifying ST25 and ST28 strains in North America (Fittipaldi et al., 2011).

Various subtyping methods, including serotyping and MSLT, have been used for evaluating the genetic diversity of *S. suis* isolates and identifying patterns specific to clinical isolates. Pulsed-field gel electrophoresis (PFGE) has been used for evaluating the genetic diversity of *S. suis* serotype 2, 1/2, 3, 7, and 9 strains (Berthelot-Hérault et al., 2002; Vela et al., 2003). Although PFGE has high discriminatory power, typing a large number of isolates is time consuming and labor intensive. Unique randomly amplified polymorphic DNA patterns have been recovered from *S. suis* isolates from diseased pigs and correlated with the production of virulence markers (Chatellier et al., 1999; Maneerat et al., 2013). However, these analyses were mainly focused on serotype 2 strains.

Multiplex PCR assays were developed for the differentiation of isolates into serotypes and detection of multiple VAGs (Silva et al., 2006; Kerdsin et al., 2014; Okura et al., 2014). A limitation of multiplex PCR assays is the number of targets that can successfully be tested in a single assay (Yang and Rothman, 2004). In this study, we utilized pan-genome analysis in conjunction with serotyping, MLST, and VAG profiling as a subtyping tool for *S. suis*. Whole genome sequencing (WGS)-based approaches, such as comparative genome hybridization, minimum core genome sequence typing, pan-genome and Bayesian analysis of population structure, and genome-wide association studies, have been used in combination with phenotypic methods for the identification and classification of *S. suis* strains into groups of differing levels of virulence (de Greeff et al., 2011; Zhang et al., 2011; Chen et al., 2013; Weinert et al., 2015; Willemse et al., 2016; Wileman et al., 2019). WGS-based approaches have multiple advantages to molecular subtyping techniques, such as the ability to characterize the entire genome, a higher discriminatory power capable of discriminating closely related strains, the ability to perform in silico (via computer simulation) analyses, and access to a vast number of bioinformatics tools for the analysis of whole genomes (Quainoo et al., 2017; Uelze et al., 2020).

Novel published VAGs for the identification of pathogenic *S. suis* isolates were selected using a chi-square test and a LASSO regression model testing associations between published VAGs and pathotype. As a result, the two genes *ofs* and *srtF* were selected as the 'best' indicators of pathogenicity for isolates in our study. The *ofs* gene encodes a serum opacity factor and was associated with virulence attenuation in an experimental pig model (Baums et al., 2006). The *srtF* gene encodes a class C sortase and is part of the *srtF* pilus gene cluster composed of four genes, *srtF*, *sipF*, *sfp1*, and *sfp2* (Fittipaldi et al., 2010). *SrtF* gene mutants of *S. suis* serotype 2 ST1 strain P1/7 caused attenuation of virulence in an intranasal caesarean-derived colostrum-deprived (CDCD) pig model (Faulds-Pain et al., 2019). However, the presence of the pilus gene cluster does not guarantee pilus protein expression (Fittipaldi et al., 2011). Our research identifies the genes as markers for pathogenicity and not the expression of proteins. The percentage of isolates containing the *ofs*+/ *srtF*+ genotype that were classified as pathogenic increased

from 79 to 96% (132/137) when excluding the possibly opportunistic pathotype (isolates possibly associated with respiratory disease) from the analysis. The proposed pathogenic genotype for predicting pathogenicity was further tested in a validation set consisting of 32 *S. suis* isolates to evaluate the likelihood of these two genes identifying pathogenic strains in other swine production companies. The *ofs*+/*srtF*+ genotype was observed in 73.7% (14/19) of the 'pathogenic' isolates, together indicating a ≥ 74% probability that an isolate will be classified as pathogenic given the proposed genotype. The proposed *ofs*+/*srtF*+ genotype, in complement to the classical VAGs for ST1, identifies pathogenic strains in the United States. A potential application of this research is the development of a diagnostic PCR test targeting these two proposed VAGs.

Nineteen of 139 isolates in the pathogenic pathotype lacked the *ofs*+/*srtF*+ genotype suggesting the possibility of misclassification of these isolates based on tissue source (systemic versus non-systemic). In addition to pathogen-specific traits, environmental and management conditions and host traits contribute to the development of *S. suis* disease. These factors include temperature fluctuations, overcrowding, concurrent infections, and host immunity and genetics (Dee et al., 1993; Reams et al., 1994; Segura et al., 2016). The *ofs*+/*srtF*+ genotype was identified in five isolates in the commensal pathotype but four of these isolates were characterized as CC1 or CC94, which are generally pathogenic subtypes (Kerdsin et al., 2018; Hatrongjit et al., 2020). The five commensal isolates are present in Cluster I (Fig. 3.1), which represents a cluster containing all three pathotypes, indicating these isolates share similar VAGs with pathogenic isolates. Virulent strains have been previously isolated from the nasal cavities and tonsils of clinically healthy pigs, so characterization by tissue source can be misleading (Brisebois et al., 1990; Marois et al., 2007).

Pan-genome analysis in combination with metadata (serotype, ST/CC, VA genotype) was used in this study as a subtyping tool to describe the genetic diversity of *S. suis* isolates within a production company and between companies for epidemiological purposes. The differentiation of *S. suis* may provide information on the origin of isolates (geographical location, year, source, etc.) or aid in the identification and tracking of

strains over time (Beaudoin et al., 1992; Doto et al., 2016; Dutkiewicz et al., 2017). Isolates from the pathogenic pathotype in this study formed distinct clusters with correlation to CC and VA genotypes, which is consistent with previous studies (Maneerat et al., 2013; Dong et al., 2017). The same CC-VA genotype patterns were identified in multiple production companies, suggesting a lack of association between production company, CC, or VA genotype. These observed patterns may be widespread as opposed to originating from a common source of infection as previously suggested (Mwaniki et al., 1994; Morales et al., 2015; Denich et al., 2020). Furthermore, the high genetic similarity and identical CC and VAG genotypes within a pan-genome cluster (such as in cluster A in production company E) are indicative of a clone, providing useful information for the identification and tracking of clones over time (Allgaier et al., 2001; van Belkum et al., 2007; Du et al., 2017). Thus, the use of WGS to complement metadata (e.g. epidemiological, clinical and demographical data) provides a valuable tool for subtyping *S. suis* as part of epidemiological studies (Rantsiou et al., 2018; Van Goethem et al., 2019). Further, pan-genome analysis of U.S. *S. suis* isolates may be used to identify candidate VAGs not yet identified or characterized.

The differentiation of *S. suis* isolates is also crucial for the development of autogenous vaccines (Rieckmann et al., 2020). Different strains have been recovered from diseased pigs from the same herd and selecting the strain or strains associated with disease is challenging (Mogollon et al., 1991; Reams et al., 1996; Martinez et al., 2002; Vela et al., 2003). For the validation set, multiple CC-VA genotype patterns were found among the 'pathogenic' clusters, indicating multiple clones were present in this production company. This diversity of isolates is supported by the identification of five serotypes (1, 1/2, 2, 14, and 7) in the validation set, all of which are generally pathogenic subtypes (Boetner et al., 1987; Smith et al., 1999; Segura et al., 2020). Despite the diversity of clinical strains in the same herd, previous reports indicate a specific strain is the predominant cause of disease and the primary candidate for an autogenous vaccine (Mogollon et al., 1991; Reams et al., 1996; Torremorell and Pijoan, 1998; Martinez et al., 2002). CC1 was predominantly identified in this production company, and these CC1/ST1 isolates (cluster e and f) demonstrated similar gene content (99% similarity)

and genotypes but had different serotypes (serotypes 1/14 vs serotype 2). These results suggest two sub-populations with differences in virulence potential and the need for multiple isolates in a vaccine (Haesebrouck et al., 2004; Hopkins et al., 2019). On the other hand, the CC28 isolates (cluster d) demonstrated similar gene content (92–99% similarity), serotypes, and genotypes, suggesting similar virulence potential, and the selection of a single isolate for vaccine (Fittipaldi et al., 2011; Athey et al., 2015). As these isolates came from the same production company, all three isolates may by recommended for vaccine development. In addition to the genetic diversity of *S. suis* isolates, historical background of a production company should be considered while selecting isolates. Historical factors such as prior on-farm identification of *S. suis*, historic and current sources of replacement animals, and other confounding disease factors can further support the inclusion of multiple isolates in vaccine development.

In this study, the current distribution of published, including classical, VAGs in U.S. isolates was determined, which indicated that classical VAGs are not sufficient to differentiate pathogenic and commensal U.S. strains. Of the 71 published VAGs investigated, the *ofs* and *srtF* genes were shown to be stronger predictors of pathogenicity in both a training and a validation set of isolates. Furthermore, a WGS-based approach was used to determine the genetic diversity of isolates demonstrating its use in epidemiological studies and vaccine isolate selection.

3.6 Acknowledgements

The authors would like to thank the bacteriology section at the University of Minnesota Veterinary Diagnostic Lab for their technical assistance.

3.7 Funding

The research project was partially funded by the Minnesota State Legislature through the Rapid Agricultural Response Fund, managed by the Minnesota Agricultural Experiment Station (project #00057268). April A. Estrada's stipend was funded by the USDA National Needs Fellowship Award, 2016–38420-25288.

3.8 Figures

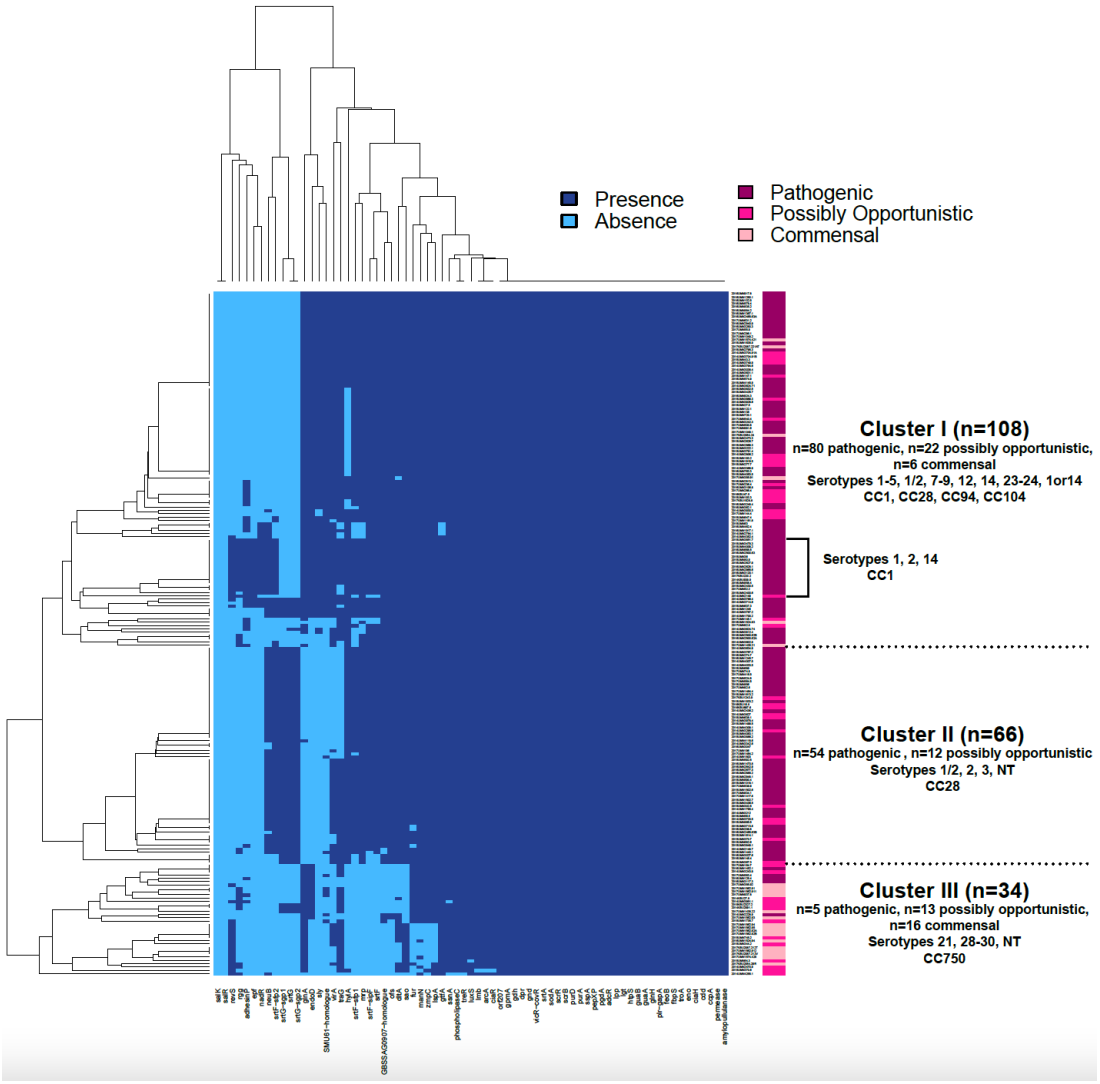

Figure 3.1 Virulence-associated gene (VAG) profiling of 208 *S. suis* isolates. Heatmap illustrating the presence and absence of 71 previously published VAGs in 208 isolates. Isolates are annotated (right) by pathotype (pathogenic, possibly opportunistic, commensal). Clustering of the 208 *S. suis* isolates by VAGs illustrated three clusters (Clusters I-III), two of which (Clusters II and III) suggest associations between VAGs and pathotype. Serotype and clonal complex (CC) distributions in each of the three clades are also denoted.

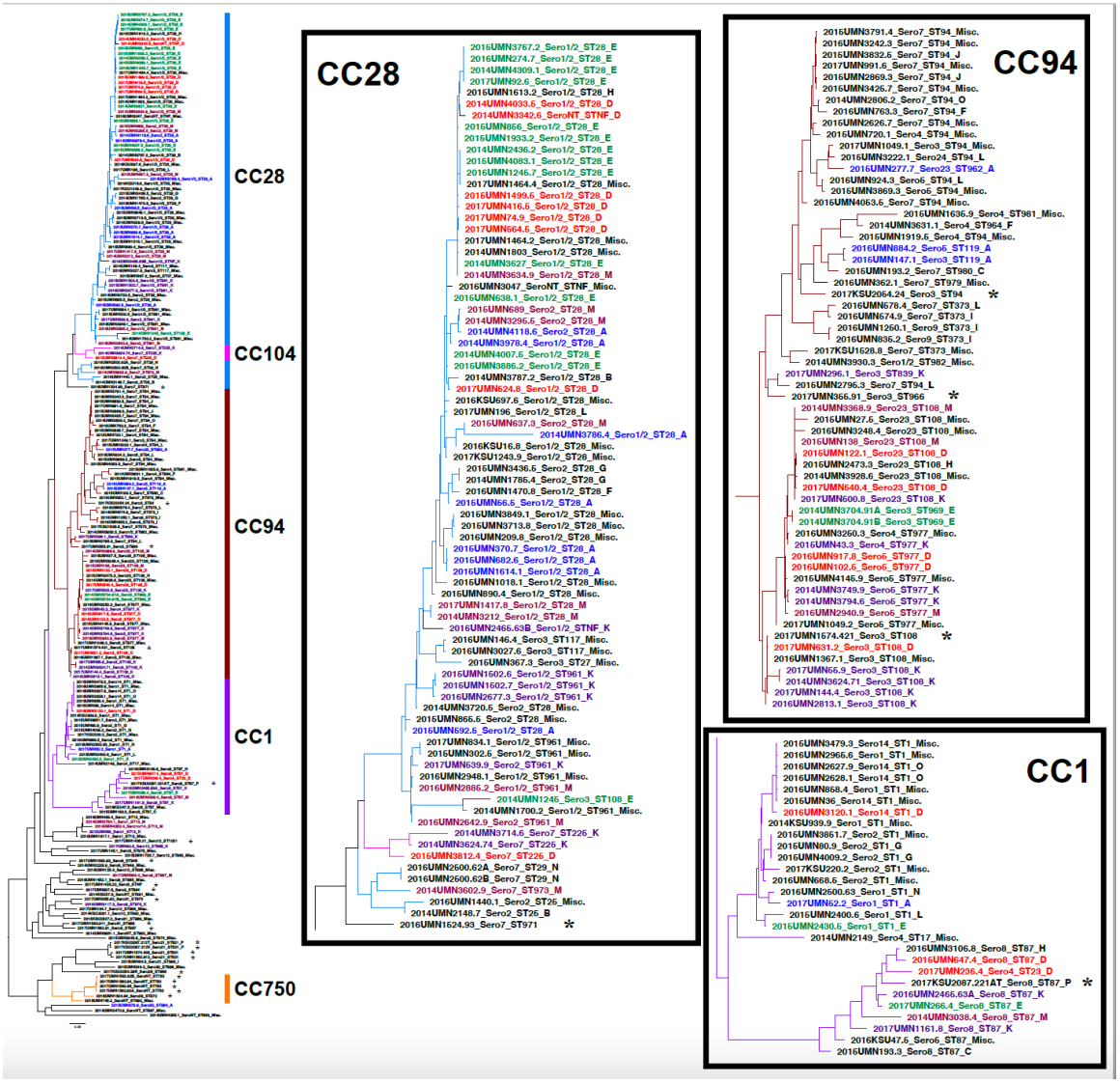

Figure 3.2 Relatedness of 208 *S. suis* isolates by pan-genome analysis. Genetic relationships between isolates are based on the presence and absence of 8373 gene clusters among 208 *S. suis* genomes. The phylogenetic tree is colored-coded (branches) and labeled (right) by CC; multiple STs did not form a CC or formed a CC without a primary founder. Isolates from at least sixteen swine production companies (A-P) (≥ 2 isolates each) were identified in the data set. Misc. refers to miscellaneous production companies (single isolates each). Isolates belonging to the five predominant production companies (A, D, E, K, and M) are color-coded by their respective production company. * strains in the commensal pathotype (n = 22).

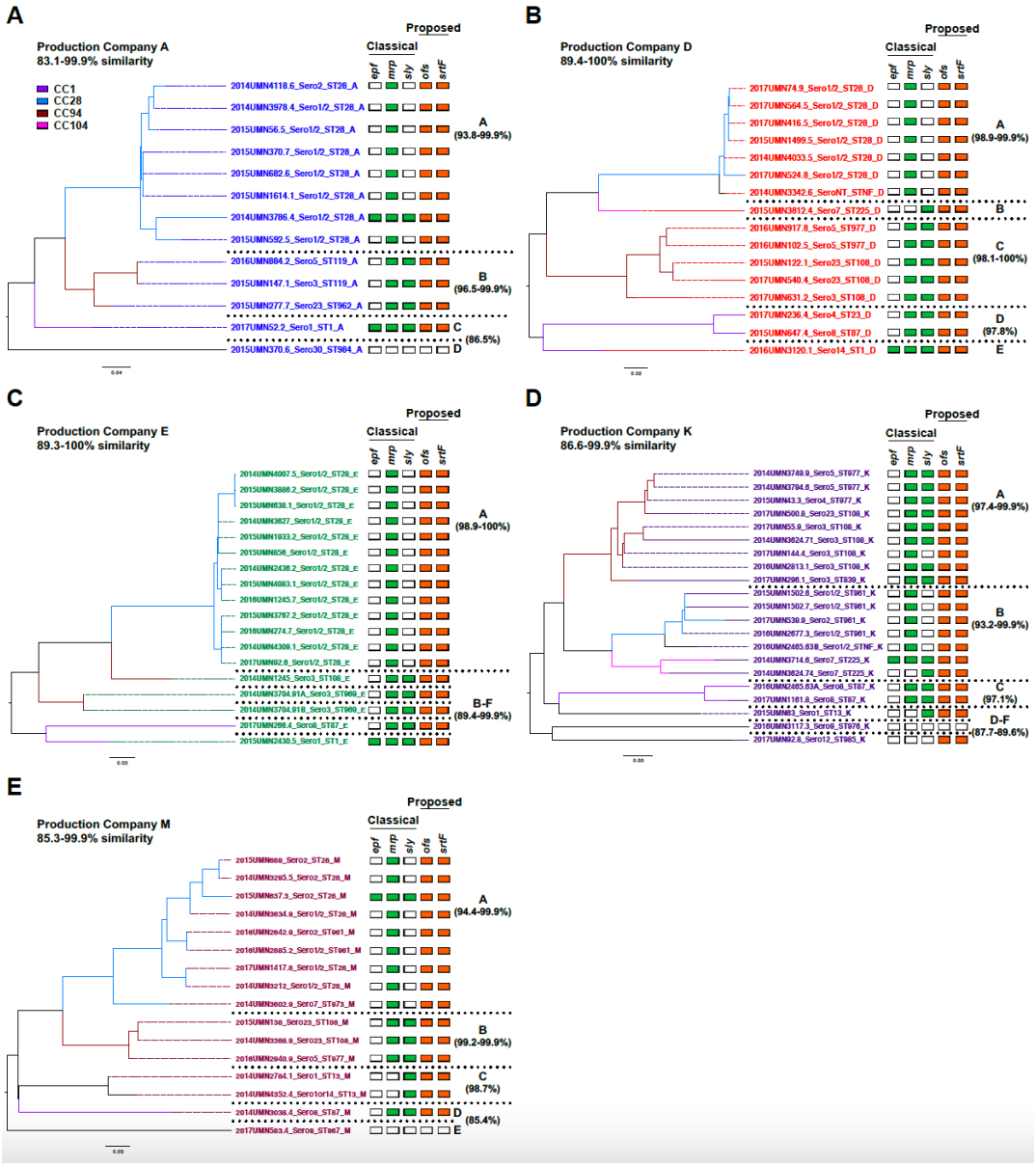

Figure 3.3 Pan-genome analysis of isolates from the five predominant production companies. The predominant production companies are presented as A, D, E, K, and M. Color-coding of isolate names by production company and color-coding of phylogenetic tree branches by CC follow the same color schemes as Figure 3.2. The percent similarity of isolates within a cluster is defined as the percentage of shared genes from a total of 8373 genes. The presence of the classical VAGs *epf*, *mrp*, and *sly* is represented in green and the proposed VAGs *ofs* and *srtF* in orange.

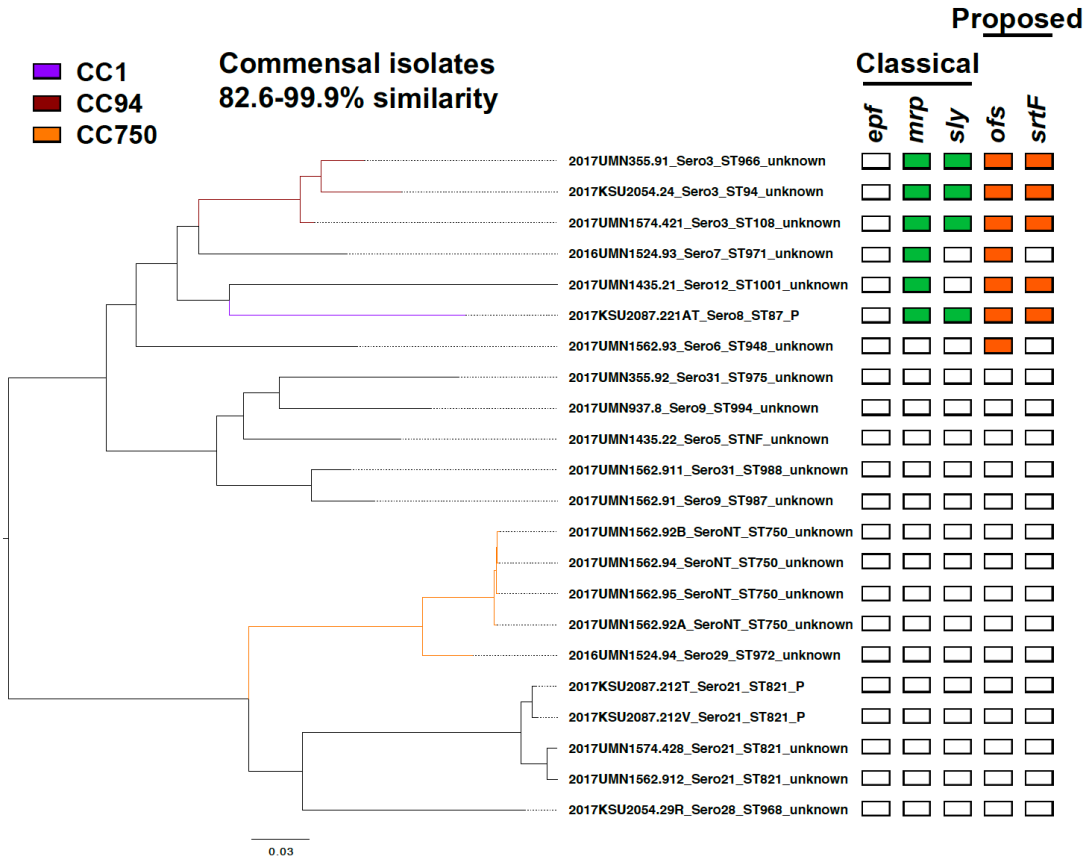

Figure 3.4 Pan-genome analysis of the 22 commensal isolates. Color-coding of phylogenetic tree branches by CC follows the same color scheme as Figure 3.2. The presence of the classical VAGs *epf*, *mrp*, and *sly* is represented in green and the proposed VAGs *ofs* and *srtF* in orange.

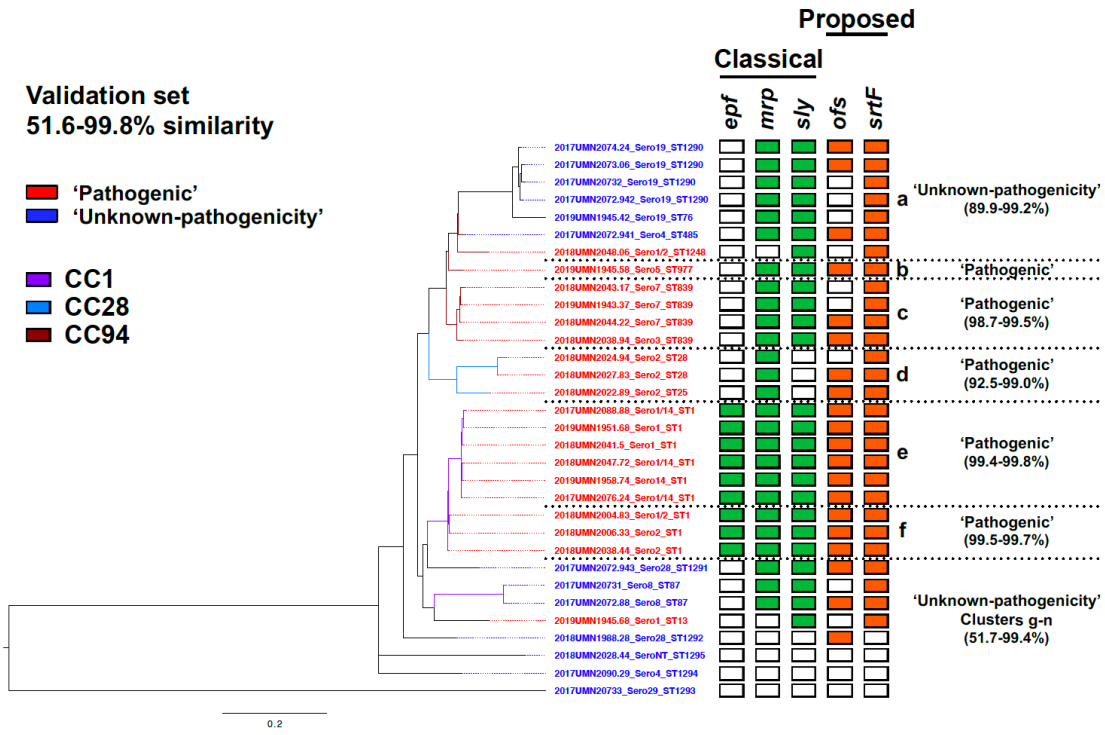

Figure 3.5 Pan-genome analysis of the validation set of 32 *S. suis* isolates. Isolates are color coded by classification, 'pathogenic' (red) or 'unknown- pathogenicity' (blue). The phylogenetic tree branches are colored-coded by CC. The percent similarity of isolates within a cluster is defined as the percentage of shared genes from a total of 7078 genes. The presence of the classical VAGs *epf*, *mrp*, and *sly* is represented in green and the proposed VAGs *ofs* and *srtF* in orange.

3.9 Tables

Table 3.1 Classical VAGs identified in the pathogenic pathotype of *S. suis* isolates (*n* = 139)

Subtypes			**Percentage of positive**		
			epf (*n* = 20, 14.4%)	**mrp (*n* = 127, 91.4%)**	**sly (*n* = 77, 55.4%)**
Serotypes		1 (*n* = 11)	35.0	5.5	14.3
		1/2 (*n* = 54)	5.0	35.4	3.9
		2 (*n* = 17)	30.0	11.0	7.8
		3 (*n* = 18)	0.0	7.9	10.4
		4 (*n* = 8)	0.0	3.1	5.2
		5 (*n* = 13)	0.0	6.3	10.4
		6 (*n* = 2)	0.0	0.0	0.0
		7 (*n* = 23)	5.0	13.4	20.8
		8 (*n* = 8)	0.0	3.1	5.2
		9 (*n* = 8)	0.0	1.6	2.6
		10 (*n* = 3)	0.0	0.0	0.0
		14 (*n* = 5)	25.0	3.9	6.5
		23 (*n* = 10)	0.0	6.3	10.4
		24 (*n* = 1)	0.0	0.8	1.3
		1or14[a] (*n* = 1)	0.0	0.0	1.3
		NT (*n* = 11)	0.0	1.6	0.0
Multilocus Sequence Type	**CC1[b]**	ST1 (*n* = 17)	85.0	13.4	22.1
	CC1	ST87 (*n* = 9)	0.0	3.1	5.2
	CC28	ST25 (*n* = 2)	0.0	1.6	0.0
	CC28	ST28 (*n* = 52)	10.0	33.1	3.9
	CC28	ST29 (*n* = 2)	0.0	1.6	0.0
	CC28	ST117 (*n* = 2)	0.0	1.6	0.0

CC28	ST961 (*n* = 10)	0.0	7.1	1.3
CC28	ST973 (*n* = 1)	0.0	0.8	0.0
CC94	ST94 (*n* = 18)	0.0	11.0	18.2
CC94	ST108 (*n* = 17)	0.0	11.0	18.2
CC94	ST119 (*n* = 2)	0.0	0.8	1.3
CC94	ST373 (*n* = 5)	0.0	3.1	5.2
CC94	ST839 (*n* = 1)	0.0	0.8	1.3
CC94	ST964 (*n* = 1)	0.0	0.8	1.3
CC94	ST977 (*n* = 9)	0.0	5.5	9.1
CC94	ST981 (*n* = 1)	0.0	0.8	1.3
CC104	ST225 (*n* = 3)	5.0	0.8	3.9
	ST13 (*n* = 5)	0.0	0.0	6.5
	ST949 (*n* = 1)	0.0	0.0	0.0
	ST965 (*n* = 1)	0.0	0.0	0.0
	ST967 (*n* = 1)	0.0	0.0	0.0
	ST976 (*n* = 1)	0.0	0.0	0.0
	ST979 (*n* = 1)	0.0	0.8	1.3
	ST995 (*n* = 1)	0.0	0.0	0.0
	NF (*n* = 4)	0.0	2.4	0.0

[a] Could not differentiate serotypes 1 and 14 by coagglutination, PCR, and WGS

[b] Clonal complexes (CCs) determined in our previous study (Estrada et al., 2019)

NT = Unresolved serotype by coagglutination, PCR, and WGS

NF = ST could not be determined

Table 3.2 Classical VAGs *epf*, *mrp*, and *sly* genotypes identified in the pathotypes of *S. suis* isolates (*n* = 208)

Virulence-associated gene or genotype	**No. possessing gene or genotype (%)**	**Pathogenic (*n* = 139)**	**Possibly Opportunistic (*n* = 47)**	**Commensal (*n* = 22)**	**ST1 (*n* = 17)**	**ST25 (*n* = 2)**	**ST28 (*n* = 52)**
epf	21 (10.1%)	20	1	0			
mrp	165 (79.3%)	127	32	6			
sly	99 (47.6%)	77	18	4			
epf−/mrp+/sly-	73 (35.1%)	57	14	2	0	2	49
epf−/mrp+/sly+	71 (34.1%)	50	17	4	0	0	1
epf−/mrp−/sly-	36 (17.3%)	5	15	16	0	0	0
epf+/mrp+/sly+	21 (10.1%)	20	1	0	17	0	2
epf−/mrp−/sly+	7 (3.4%)	7	0	0	0	0	0
epf+/mrp+/sly-	0	0	0	0	0	0	0
epf+/mrp−/sly+	0	0	0	0	0	0	0
epf+/mrp−/sly-	0	0	0	0	0	0	0

Table 3.3 Distribution of 71 VAGs for the pathotypes of *S. suis* isolates (*n* = 208)

VAG	No. containing the VAG	Pathogenic (*n* = 139)	Possibly Opportunistic (*n* = 47)	Commensal (*n* = 22)
adcR[a]	208	139	47	22
amylopullulanase[a]	208	139	47	22
ccpA[a]	208	139	47	22
cdd[a]	208	139	47	22
ciaH[a]	208	139	47	22
dpr[a]	208	139	47	22
eno[a]	208	139	47	22
fbpS[a]	208	139	47	22
feoB[a]	208	139	47	22
gdh[a]	208	139	47	22
glnH[a]	208	139	47	22
gnd[a]	208	139	47	22
gpmA[a]	208	139	47	22
guaA[a]	208	139	47	22
guaB[a]	208	139	47	22
htpS[a]	208	139	47	22
lgt[a]	208	139	47	22
lpp[a]	208	139	47	22
orf207[a]	208	139	47	22
pepXP[a]	208	139	47	22
permease[a]	208	139	47	22
pgdA[a]	208	139	47	22
plr-gapA[a]	208	139	47	22
purA[a]	208	139	47	22
purD[a]	208	139	47	22
scrB[a]	208	139	47	22
scrR[a]	208	139	47	22
sodA[a]	208	139	47	22
srtA[a]	208	139	47	22
sspA[a]	208	139	47	22

troA[a]	208	139	47	22
vicR-covR[a]	208	139	47	22
arcA	207	139	46	22
ciaR	207	139	46	22
lmb	207	139	46	22
luxS	207	139	46	22
phospholipaseC	207	139	46	22
treR	207	139	46	22
ssnA	206	139	45	22
lspA	196	139	44	13
manN	196	139	43	14
zmpC	193	139	42	12
fur	190	136	41	13
gtfA	204	135	47	22
dltA[b]	186	135	39	12
ofs[b]	176	135	34	7
sao	189	134	39	16
GBSSAG0907-homologue[b]	175	133	35	7
srtF[b]	168	132	31	5
srtF-sipF[b]	167	131	31	5
mrp[b]	165	127	32	6
srtF-sfp1[b]	155	120	31	4
hylA[b]	145	112	28	5
virA[b]	142	105	26	11
traG[b]	158	100	39	19
glnA[b]	138	83	34	21
endoD[b]	116	80	26	10
SMU61-homologue[b]	107	78	22	7
sly[b]	99	77	18	4
neuB[bc]	90	77	13	0

srtF-sfp2[b]	90	77	12	1
srtG	76	61	14	1
srtG-sgp2	76	61	14	1
srtG-sgp1	75	60	14	1
adhesinP	43	30	10	3
nadR[c]	25	25	0	0
rgg	33	21	7	5
revS	40	20	9	11
epf [c]	21	20	1	0
salK[c]	0	0	0	0
salR[c]	0	0	0	0

[a] Represents VAGs identified in all the isolates

[b] Represents VAGs tested by chi-square

[c] Represents lack of VAGs in the commensal pathotype

Table 3.4 LASSO results for the four candidate VAGs in the pathotypes of *S. suis* isolates (n = 208)

VAG(s)	No. containing the VAG(s)	Pathogenic (n = 139)		Possibly Opportunistic (n = 47)		Commensal (n = 22)	
		No.	Proportion[a]	No.	Proportion[a]	No.	Proportion[a]
ofs	176	135	0.767	34	0.193	7	0.040
srtF	168	132	0.786	31	0.185	5	0.030
neuB	90	77	0.856	13	0.144	0	0.000
srtF-sfp2	90	77	0.856	12	0.133	1	0.011
ofs and srtF	168	132	0.786	31	0.185	5	0.030

[a] positive isolates in the pathotype divided by the number of isolates containing the VAG(s)

CHAPTER 4

Comparative analysis of *Streptococcus suis* genomes identify novel virulence-associated genes in U.S. isolates

April A. Estrada[1*], Marcelo Gottschalk[2], Douglas G. Marthaler[3], Connie J. Gebhart[1]

[1]Department of Veterinary and Biomedical Sciences, College of Veterinary Medicine, University of Minnesota, Saint Paul, MN, USA

[2]Faculty of Veterinary Medicine, University of Montreal, Saint-Hyacinthe, Quebec, Canada

[3]Kansas City, KS, USA, marth027@umn.edu

4.1 Synopsis

Streptococcus suis is a significant economic and welfare concern in the swine industry. Classical molecular microbiology techniques have been utilized to traditionally characterize *S. suis* strains. Pan-genome analysis provides an in-silico approach for the discovery of genes involved in pathogenesis in bacterial pathogens. In this study, we performed pan-genome analysis of 208 *S. suis* isolates classified into the pathogenic, possibly opportunistic, and commensal pathotypes to identify novel candidate virulence-associated genes (VAGs) of *S. suis*. Using the chi-square test and LASSO regression model, three accessory pan-genes corresponding to *S. suis* strain P1/7 markers *SSU_RS09155*, *SSU_RS09525*, and *SSU_RS03100* (> 95% identity) were identified as having a significant association with the pathogenic pathotype. The proposed novel *SSU_RS09155+*/*SSU_RS09525+*/*SSU_RS03100+* genotype identified 96% of the pathogenic pathotype suggesting a novel genotyping scheme for predicting the pathogenicity of *S. suis* isolates in the United States. In addition, mobile genetic elements carrying antimicrobial resistance genes (ARGs) and VAGs were identified but did not appear to play a major role in the spread of ARGs and VAGs.

Impact statement The use of whole genome sequencing along with clinical metadata for the investigation of *S. suis* disease have revealed important aspects of the diversity and pathogenesis of this species. For example, whole genome sequencing-based approaches have been instrumental in determining the genomic characteristics of virulent European and Asian strains and the development of pathotyping tools. However, there is limited information on genomic characterization of U.S. isolates, and it is unclear whether the findings from Europe or Asia can be extrapolated to North American swine herds. In this study, we investigated the pan-genome, novel candidate virulence-associated genes, antimicrobial resistance genes (ARGs), and mobile genetic elements (MGEs) of a large collection of U.S. isolates. We demonstrated that isolates belonging to the three pathotypes share ARGs, MGE-linked genes, and various other genes, and can be differentiated by the presence of three novel pan-genome based markers. The findings of

this work may emphasize the advantages of whole genome sequencing of U.S. *S. suis* isolates and the need for contemporary comparative studies.

Keywords *Streptococcus suis*, virulence-associated genes (VAGs), pathotype, comparative genomics

4.2 Introduction

Streptococcus suis is a significant economic and welfare concern in the swine industry as pathogenic strains cause a range of clinical signs including meningitis, arthritis, endocarditis, and septicemia (Gottschalk and Segura, 2019). Commensal strains naturally colonize the upper respiratory tract of pigs without causing clinical disease (Arends et al., 1984). Characterization techniques capable of identifying strains of clinical significance are important for the control of *S. suis* disease. Virulence-associated factors contributing to the virulence of *S. suis* strains (mainly for serotype 2) have been described but are not consistently present in clinical isolates (Fittipaldi et al., 2012; Segura et al., 2017). *S. suis* virulence-associated factors remain unknown or are geographically specific.

Increasing availability of next generation sequencing technologies and generation of large amounts of data led to the development of numerous programs and software tools for bacterial typing (Carriço et al., 2013; Edwards and Holt, 2013). Pan-genome analysis characterizes the diversity within a bacterial species as it describes core traits shared by all strains and unique/accessory traits shared only by some strains (Medini et al., 2005). *S. suis* possesses an open pan-genome in which the number of unique or accessory genes increases as more genomes are sequenced (Zhang et al., 2011). The open pan-genome contributes to high species diversity and is typically indicative of high rates of horizontal gene transfer by mobile genetic elements (MGEs) (Medini et al., 2005; Rouli et al., 2015). Functional analyses of the *S. suis* pan-genome revealed differences in the functional annotation of the core genome compared to the accessory genome (Zhang et al., 2011; Guo et al., 2020), supporting the distinct roles of these two components in explaining genomic diversity. Comparative genomics can also be used as a subtyping tool

for *S. suis*. Serotype 1, 2, 7, and 9 strains were differentiated by comparative genome hybridization or bayesian analysis of population structure analysis into multiple genomic groups, some of which correlated to virulence and traditional molecular subtyping techniques (de Greeff et al., 2011; Willemse et al., 2016).

The ability to identify virulence determinants within the accessory genome has been demonstrated for various bacterial pathogens such as *Escherichia coli*, *Pseudomonas aeruginosa*, and *Streptococcus* spp. (Rasko et al., 2008; Kilian and Tettelin, 2019; Allen et al., 2020). Known and putative VAGs were over-represented in systemic isolates of *S. suis* compared to respiratory and non-clinical isolates (Weinert et al., 2015). In a separate study, a genome-wide association study was utilized to identify three novel genetic markers for differentiating *S. suis* isolates into invasive disease-associated and non-disease-associated groups (Wileman et al., 2019). That pathotyping tool, which consisted of a multiplex-PCR assay, demonstrated performance metrics (e.g., sensitivity, specificity) similar to serotyping, multilocus sequence typing (MLST), VAG profiling, and minimum core genome sequence typing. Two recent publications on the comparative analysis of virulent and avirulent strains identified two different sets of novel virulence-associated genes, supporting previous claims of an abundance of uncharacterized virulence determinants in the accessory genome of pathogenic strains (Guo et al., 2020; Nicholson et al., 2020).

Mobile genetic elements facilitate intracellular and intercellular movement of genes and contribute to the evolution of pathogenic bacteria such as *S. suis* (Frost et al., 2005; Huang et al., 2016; Partridge et al., 2018). MGEs are a large group of DNA fragments that include plasmids, integrative conjugative elements (ICE), integrative mobilizable elements (IME), composite transposons, and pathogenicity islands (PAI), some of which are mobile by conjugation. Proteins encoded by MGEs have various functions that can be described as core traits required for replication and advantageous/adaptive traits (e.g., antimicrobial resistance genes (ARGs), virulence-associated genes (VAGs)) (Frost et al., 2005; Rankin et al., 2011). Previous studies demonstrated the transfer of conjugative ICESsu32457 to *S. agalactiae*, *S. pneumoniae*,

and *S. pyogenes*, indicating the dissemination of resistance genes among closely related species (Palmieri et al., 2012; Marini et al., 2015). Virulence-associated factors associated with mobile genetic elements have been identified in *E. coli*, *Legionella pneumophila*, *P. aeruginosa* and *S. suis* (Rasko et al., 2008; D'Auria et al., 2010; Zheng et al., 2011; Allen et al., 2020). An 89kb PAI was first identified in the highly virulent and zoonotic Chinese *S. suis* strain 05ZYH33, and it contains both resistance genes and putative VAGs (Chen et al., 2007; Li et al., 2008). There is evidence that *S. suis* serves as an MGE reservoir for other streptococci, and there are many poorly characterized MGEs yet to be discovered (Huang et al., 2016; Hadjirin et al., 2020).

Comparative genomics provides a powerful tool for the characterization and subtyping of *S. suis*. However, a contemporary comparative pan-genomic study targeting U.S. *S. suis* strains is still incomplete. We performed comparative genome analysis on a set of 208 *S. suis* isolates from North America (mainly the United States) to identify accessory genes corresponding to the pathogenic pathotype and thus may serve as novel candidate virulence-associated genes of *S. suis*. The identification of candidate VAGs may elucidate a novel VAG genotyping scheme for predicting the pathogenicity of *S. suis* isolates in the United States. Further, we determined the diversity of MGEs in U.S. *S. suis* isolates and investigated their role in the dissemination of ARGs and VAGs among *S. suis* pathotypes.

4.3 Materials and methods

4.3.1 Sample selection and genome assembly

The study utilized all 208 *S. suis* genomes (referred to as the training set) previously described by Estrada et al. (Estrada et al., 2019). These isolates were previously classified into pathotypes and characterized by serotyping, MLST, and VAG profiling. Briefly the three pathotypes are pathogenic (systemic tissues), possibly opportunistic (lung tissues), and commensal (laryngeal, tonsil, or nasal samples). Genome assembly was performed on Illumina sequencing data of the 208 *S. suis* isolates SRA accession numbers SRR9123061-SRR9123268. Genome assemblies (contigs) were

generated using the SKESA *de-novo* assembler (v2.4.0) (Souvorov et al., 2018) with default kmer settings. QUAST (v4.5) (Gurevich et al., 2013) was used to evaluate the genome assemblies and generate summary statistics (e.g. genome length, GC content, N50) for contigs ≥ 500 bp. Only contigs ≥ 500 bp were kept for annotation by Prokka (v1.14.6) (Seemann, 2014) to predict coding sequences. The pan-genome was annotated using Roary (v3.13.0) (Page et al., 2015) using a 90% BLASTp identity cut-off to define clusters of genes and allowing paralog clustering. Gene clusters present in 99% (≥ 206/208) of genomes were classified as core genes. Two different Roary analyses were performed. The first analysis utilized all 208 genomes while the second analysis utilized the 161 genomes representing only the pathogenic and commensal pathotypes.

4.3.2 Functional annotation

The Clusters of Orthologous Groups of proteins (COG) database (2014) (Galperin et al., 2015) was utilized to predict protein functions. For each gene cluster identified by Roary, a representative protein sequence was selected, and BLASTp searches against the COG database were performed. COG functional classes from searches meeting the thresholds (coverage ≥ 70%, identity ≥ 70%, and e-value ≤ 10-5) were plotted using R software (v3.6.1) (R Core Team, 2017).

4.3.3 Statistical analysis

Basic data transformation and statistical analyses were performed using R software. Associations between *S. suis* accessory genes and pathotype were investigated. The data were filtered by removing genes detected in less than 50% (≤ 69/139) of isolates within the pathogenic pathotype and detected in more than 50% (≥ 11/22) of isolates within the commensal pathotype. Remaining accessory genes were individually tested by chi-square using a 3x2 table comparing the three pathotypes and the status (presence/absence) of individual genes. Genes lacking a significant (p-value < 0.05) association with pathotype were removed from the analysis. The remaining genes were analyzed using the LASSO (Least Absolute Shrinkage and Selection Operator) shrinkage regression model.

4.3.4 LASSO shrinkage regression model

A LASSO shrinkage regression model was used, as previously described by Estrada et al (Estrada et al., 2021), to determine the fewest number of accessory genes that may serve as predictors of pathogenicity; the pathogenic pathotype served as the indicator of pathogenicity. The LASSO analysis (100 iterations) was performed on each Roary data set, the first data set consisting of all 208 isolates, and the second consisting of a subset of 161 isolates (Figure 1). Genes identified in both LASSO analyses were selected as the 'best' predictors of pathogenicity.

4.3.5 Identification of antimicrobial resistance genes

The presence of antimicrobial resistance genes (ARGs) was predicted in all 208 *S. suis* draft genomes (contigs) using the Comprehensive Antibiotic Resistance Database (CARD) (v3.1.0) and the CARD BLAST command-line program (Alcock et al., 2020) (BLASTn, ≥ 90% identity and ≥ 60% coverage).

4.3.6 Identification of mobile genetic elements

Plasmids and other genetic elements (insertion sequences, ICE, IME, composite transposons, etc.) were identified in the *S. suis* draft genomes using the command-line PlasmidFinder (Carattoli et al., 2014) and MobileElementFinder (Johansson et al., 2021) programs (default parameters: ≥ 90% identity and ≥ 60% coverage). The PlasmidFinder database, which is a curated database of plasmid replicons, was updated to include *S. suis* plasmid replicons (pNSUI060a: CP012912, pNSUI060b: CP012913, HN105 unnamed plasmid1: CP029399, pSRD478: CP017089, pISU2812: CP017093, pISU2514: CP030021, pISU2614: CP031378, pYSJ17: CP032065). The MobileElementFinder database contains genetic elements from several public databases including RefSeq, Tn registry and ICEberg. The presence of the *S. suis* strain 05ZYH33 89K candidate pathogenicity island (PAI) was determined *in-silico* by screening for the CH1/CH2, CH3/CH4 and CH5/CH6 DNA sequences (Chen et al., 2007; Schmid et al., 2011) using the SRST2 (Short Read Sequence Typing for Bacterial Pathogens) program (≥ 90%

coverage and ≥ 90% sequence identity) (Inouye et al., 2014). The CH3 and CH4 sequence is a 716bp fragment unique to the strain 05ZYH33 89K PAI, thus isolates lacking this segment were considered negative for the PAI. PAI draft sequences were extracted from the genomes by mapping the trimmed reads to the *S. suis* strain 05ZYH33 89K PAI (Yoon et al., 2015). The plasmid, ICE, IME, composite transposon, and 89K PAI draft sequences were annotated using Prokka. Associations between the presence of MGEs and pathotype were tested by chi-square.

4.3.7 Identification of MGE-associated ARGs and VAGs

Comparative analysis to identify genetic elements carrying ARGs and putative VAGs was performed by BLAST searching MGE draft sequences against a custom database [BLASTn (dc-megablast), ≥ 80% identity and ≥ 60% coverage]. The CARD (v3.1.0), CGE ResFinder (02-19-2021), and ARG-ANNOT ARG (v6) databases, and the VF (03-01-2021), CGE VirulenceFinder (05-29-2020), and *S. suis* VAG databases were combined into one custom ARG-VAG BLAST database. Potential ARGs and VAGs were also determined from prokka annotations.

4.4 Results

Identification of core and accessory gene content

In the current study, the pan-genome (core and accessory genes) of the training set was determined. The genome lengths ranged from 1.95-2.45 Mb with an average coverage of 65-324x across the assembly and an average GC content of 40.8-41.6% (Appendix 4.1). The number of predicted protein coding sequences ranged from 1,854 to 2,399, with an average of 2,078. A total of 8,231 gene clusters were identified (Additional file 4.1). Of these, 1,189 gene clusters were classified as core genes and were present in all or nearly all genomes (≥ 206/208). A decrease in the number of conserved or core genes was observed as more genomes were added to the analysis (Figure 4.2A) while an increase in the number of unique genes was observed (Figure 4.2B). Furthermore, the total number of genes in the pan-genome continued to increase with

each additional genome (Figure 4.2C), suggesting an open pan-genome for *S. suis* and indicating a potential to discover novel genes with the sequencing of more *S. suis* strains. These genome lengths, CG content, and number of predicted proteins sequences are consistent with the *S. suis* reference strains already in the GenBank/EMBL/DDBJ database.

Cluster of Orthologous Groups classification

The COG functional classes were predicted for 98.7% (8123/8231) of gene clusters identified in the pan-genome of *S. suis.* The comparison of COG functional classes between core and accessory genes was performed only on the 995 classifications that met the criteria (coverage ≥ 70%, percent identity ≥ 70%, and e-value ≤ 10-5) (Figure 4.3A, Additional file 4.2). Core genes were more likely to be classified into functional classes T (signal transduction mechanisms), J (translation, ribosomal structure and biogenesis), F (Nucleotide transport and metabolism), and O (Posttranslational modification, protein turnover, chaperones), in decreasing order. More accessory genes were classified into classes X (Mobilome: prophages, transposons) and V (Defense mechanisms), in decreasing order (Figure 4.3B). Major differences were lacking in the distribution of COG classes by pathotype (Figure 4.3C).

Candidate VAGs

Statistical analyses were performed to test for associations between accessory genes and the three pathotypes. Of the 7,042-accessory pan-genes identified for the training set, 231 pan-genes met the criteria (p-value < 0.05, <50% commensal, >50% pathogenic, Figure 4.1A) to be further analyzed by the LASSO model. Similar analysis was performed on the commensal and pathogenic pathotypes only (n=161) (Additional file 4.3), and 223 pan-genes were associated with the pathogenic pathotype and further analyzed by the LASSO model (Figure 4.1B). Four genes corresponding to *S. suis* strain P1/7 markers *SSU_RS09155*, *SSU_RS01590*, *SSU_RS09525*, and *SSU_RS03100* (> 95% identity) were identified in both LASSO analyses (Table 4.1) and were further investigated as novel predictors of pathotype. The

SSU_RS09155+/*SSU_RS09525*+/*SSU_RS03100*+ genotype was observed in 96.4% (134/139) and 13.6% (3/22) of the pathogenic and commensal pathotypes, respectively (Table 4.2). Genotypes containing marker *SSU_RS01590* identified fewer pathogenic isolates. Thus, only markers *SSU_RS09155*, *SSU_RS09525*, and *SSU_RS03100* were selected as the 'best' predictors of pathotype. We further investigated these markers in eight virulent Eurasian reference strains by BLASTn (Table 4.3). The three markers were identified in all eight strains with ≥ 96% identity and 100% coverage.

Identification of ARGs

The presence of ARGs was predicted in all *S. suis* draft assemblies of the training set (Additional file 4.4). Fifteen ARGs representing five drug classes (aminoglycoside, lincosamide, macrolide, nucleoside, and tetracycline) were identified in at least one isolate (Figure 4.4A). The predominant ARGs were *tet*(*O*) (90.4%, 188/208) and *erm*(*B*) (69.7%, 145/208), which confer resistance to tetracycline and MLS (macrolide/lincosamide/streptogramin) antibiotics, respectively. Notably, ~77% of the commensal and possibly opportunistic pathotypes possessed macrolide resistance genes and 14.9% of the possibly opportunistic possessed aminoglycoside resistance genes compared to 66.2% and 2.9% of the pathogenic pathotype, respectively (Figure 4.4B). Multidrug resistance (ARGs conferring resistance to ≥ 3 drug classes) was determined in 11.0% (23/208) of isolates.

Identification of S. suis MGEs and MGE-associated ARGs and VAGs

The presence of various MGE types was determined *in-silico*. Plasmid replicons were predicted in 58.2% (121/208) of *S. suis* genomes (Figure 4.5A), and 15.7% (19/121) of these contained multiple (2-3) plasmid replicons. Ten different plasmids were predicted, and the predominant plasmid types were the *S. suis* plasmids pNSUI060a (54.3%, 113/208), pISU2614 (6.7%, 14/208), and pSSU1 (AB019522) (6.7%, 14/208) (Additional file 4.4). There was no association (p-value > 0.05) between the presence of plasmids and pathotype; however, a majority (61.9%, 86/139) of the pathogenic pathotype contained at least one plasmid compared to a minority (40.9%, 9/22) of the

commensal pathotype. Insertion sequences, ICEs, IMEs, and composite transposons were predicted in 98.1% (204/208), 2.4% (5/208), 1.9% (4/208), and 34.6% (72/208) of genomes, respectively. There was a significant association (p-value < 0.05) between the presence of predicted composite transposons and pathotype with 40.3% (56/139) of the pathogenic pathotype possessing at least one composite transposon compared to 9.1% (2/22) of the commensal pathotype. The *S. suis* strain 05ZYH33 89K PAI was determined by the CH3/CH4 internal PAI sequence, which was lacking in all isolates of the training set (Figure 4.5A). The CH1/CH2 (5' flanking region) and CH5/CH6 (3' flanking region) sequences were identified in 20.2% (42/208) and 39.4% (82/208) of isolates, respectively.

The ARGs and VAGs carried on the predicted MGEs were identified by a BLAST search of the MGE draft sequences to the custom ARG-VAG database (Additional file 4.4). In the training set, 5.8% (12/208) of the isolates contained ARGs aminoglycoside 6-adenylyltansferase (*SSU05_0957*) and tetracycline resistance protein (*SSU05_0922*) in the *S. suis* strain 05ZYH33 89K PAI (Figure 4.5B). Higher percentages of possibly opportunistic (12.8%, 6/47) and commensal (9.1%, 2/22) isolates carried ARGs compared to pathogenic isolates (2.9%, 4/139). ARGs (*erm*(*B*), *lnuC*, and/or *tet*(*M*)) were identified in plasmids, ICEs, and composite transposons of 1.4% (3/208) of isolates and in IMEs of 1.9% (4/208) of isolates. None of the isolates possessing these ARG-associated MGEs were classified as the commensal pathotype.

Putative VAGs were identified on PAI-like sequences of 94.2% (196/208) isolates (Figure 4.5C). These VAGs included the *S. suis* strain 05ZYH33 agglutinin receptor (*SSU05_0965*) and type IV secretory system VirB4/VirD4 components (*SSU05_0969*/*SSU05_0973*), which were present in ≥ 75% (≥ 157/208) of genomes (Additional file 4.4). Other PAI-associated VAGs present in the genomes were NisK/NisR (*SSU05_0906*/*SSU05_0907*), putative zeta toxin (*SSU05_0936*), and putative abortive infection protein (*SSU05_0966*). These PAI-associated VAGs were identified in all three pathotypes with no major difference in distribution by pathotype. VAGs were identified in less than 5% (n < 10) isolates with predicted plasmids, ICEs, IMEs, and

transposons, and included *S. suis orf207*, *revS*, and *traG*. The proposed novel VAGs *SSU_RS09155*, *SSU_RS09525*, and *SSU_RS03100* were not associated with MGEs predicted in this study, as determined by BLAST. In summary, ARGs and VAGs were identified in multiple MGE types and PAI-like regions were the most diverse.

4.5 Discussion

In this study, comparative analysis of 208 previously characterized *S. suis* isolates (Estrada et al., 2019) was performed to gain insights into the distribution and function of the *S. suis* core and accessory genomes. Functional comparisons illustrated differences in COG classes with the potential enrichment of virulence-associated genes in the accessory genome. Markers *SSU_RS09155*, *SSU_RS09525*, and *SSU_RS03100* demonstrated strong associations with the pathogenic pathotype presenting a novel candidate VAGs for identifying pathogenic *S. suis* strains in the United States. We investigated the distribution of MGEs and determined that MGEs have the potential to spread resistance genes and putative VAGs.

Functional annotation of the pan-genome was performed using the COG protein database to investigate differences in the abundance of classes between core and accessory genes. Genes involved in nucleotide transport, translation, posttranslation modifications, and signal transduction mechanisms (COG F, J, O, and T) were over-represented among the core genes. These functions (represented by COG F, J, O, and T) can more broadly be described as cellular processing and signaling, information storage, and metabolism and are responsible for basic cell function ('housekeeping') (Zheng et al., 2011; Shelyakin et al., 2019). Accessory genes were more likely to be involved in defense mechanisms and the mobilome (COG X and V), which are functions associated with host- and environmental-interactions, horizontal gene transfer, and niche-adaptation in bacterial pathogens (Ward et al., 2009; Brambila-Tapia et al., 2014; Shelyakin et al., 2019). Thus, the distribution and function of the *S. suis* pan-genome support our hypothesis that the accessory genome would be enriched for genes linked to pathogenicity.

Candidate novel VAGs in the *S. suis* accessory genome were selected using a chi-square and LASSO regression models testing associations between accessory genes and pathotype. Three pan-genes corresponding to *S. suis* strain P1/7 genes *SSU_RS09155*, *SSU_RS09525*, and *SSU_RS03100* were selected as the 'best' indicators of pathogenicity for isolates in our study. In our previous study, we demonstrated that a genotype consisting of classical *S. suis* VAGs extracellular protein factor, muramidase-released protein, and suilysin (*epf*+/*mrp*+/*sly*+) only identified 14% of the pathogenic pathotype, while a novel proposed genotype of published VAGs *ofs* and *srtF* (*ofs*+/*srtF*+) was able to identify 95% (Estrada et al., 2021). Yet, virulent strains lacking *ofs* and/or *srtF* have been reported in North America, indicating these VAGs are not essential for virulence (Fittipaldi et al., 2011, 2012). The *SSU_RS09155*+/*SSU_RS09525*+/*SSU_RS03100*+ genotype identified 96% of the pathogenic pathotype and only 14% of the commensal pathotype. Previous studies suggest Eurasian and North American strains are genotypically and phenotypically different and may possess different virulence markers (Fittipaldi et al., 2011; Gottschalk and Segura, 2019). While Eurasian type strains (serotype 2) are not predominant in the United States, these strains are still identified (Estrada et al., 2019). The identification of the *SSU_RS09155*, *SSU_RS09525*, and *SSU_RS03100* markers in all eight Eurasian strains tested suggests these markers can potentially be applied globally. The application of these virulence markers to Eurasian strains needs to be verified using a large collection of isolates including non-serotype 2 and commensal strains.

Using a genomic approach, the predicted function of these candidates VAGs and their potential relevance to *S. suis* disease were investigated. *SSU_RS09155* and *SSU_RS03100* were annotated as hypothetical proteins and have been reported in virulent *S. suis* strains such as P1/7, SC84, and GZ1. Markers *SSU_RS09155* and *SSU_RS03100* could not be further characterized by COG or by searching the NCBI protein databases, reinforcing that much of the *S. suis* genome is uncharacterized or poorly characterized. *SSU_RS09525* was annotated as an RNA-binding protein (RBP), which are involved in

post-transcriptional regulation via regulation of translation initiation, stability, and transcript elongation (Van Assche et al., 2015; Oliveira et al., 2017). RBPs are well-studied in *E. coli* and *Salmonella enterica* serovar Typhimurium and were shown to affect virulence gene expression (Liu and Romeo, 1997; Westermann et al., 2019). *SSU_RS01590* encodes a putative energy-coupling factor transporter substrate-binding protein (ECF transporter S component). ECF transporters are responsible for vitamin uptake and are essential for growth and survival, contributing to the virulence of various gram-positive bacterial pathogens (Eitinger et al., 2011; Bousis et al., 2019). The strong associations with the pathogenic pathotype and potential virulence-related functions suggest the proposed markers contribute to the pathogenicity of *S. suis*.

Fifteen ARGs conferring resistance to aminoglycoside, lincosamide, macrolide, nucleoside, and tetracycline antibiotics were identified in the draft genomes of the training set. High resistance to tetracyclines (93%, predominantly *tet*(*O*)) and erythromycin (70%, *erm*(*B*)) was observed similar to previous reports of resistance in *S. suis* in North America, Asia, and Europe (Varela et al., 2013; Seitz et al., 2016). Previous studies indicate a higher prevalence of antibiotic resistance among commensal strains, which may act as reservoirs for resistance genes (Hernandez-Garcia et al., 2017; Lugsomya et al., 2018). There was no major difference in the distribution of resistance genes by pathotype, but the commensal and possibly opportunistic pathotypes did tend to have more resistance genes.

Horizontal gene transfer of MGEs is one mechanism by which *S. suis* acquires and spreads resistance genes and putative VAGs. Thus, a genomic approach was used for the *in-silico* prediction of MGEs in *S. suis* genomes and the investigation of genes carried on these elements. Resistance MGEs were identified in only 6% of isolates, largely classified as pathogenic and possibly opportunistic (83%) and were predominantly present in PAI-like regions. However, only a few isolates (3.4%, n=7) had multiple resistance MGEs. ICE and IME are commonly found in *Streptococcus* genomes and play a major role in the dissemination of ARGs in *S. suis* (Holden et al., 2009; Libante et al., 2019). In our study, only the 69kb *S. suis* ICESsuZJ20091101-1-like ICEs (n=3) and

1.7kb *Streptococcus agalactiae* MTnSag1-like IMEs (n=4) carried resistance genes (*erm*(*B*), *lnu*C, and *tet*(*O*)). Less than 2% of isolates possessed resistance plasmids although this may not be uncommon as there are limited reports of *S. suis* plasmids carrying ARGs (Cantin et al., 1992; Wang et al., 2013) and resistance plasmids are largely found in gram-negative bacteria (Vrancianu et al., 2020). Overall, this MGE mechanism of antimicrobial resistance represents 77%, 2%, and 5% of the predicted aminoglycoside, tetracycline, and lincosamide/macrolide resistance, respectively, regardless of pathotype. Our findings suggest MGEs continue to play a role, although limited, in the spread of antimicrobial resistance.

Putative VAGs were identified among the various MGE types, but mostly on PAI-like sequences, and the predominant VAGs encode an agglutinin receptor and type IV secretion system components (T4SS). Agglutinin receptors (adhesion proteins) and T4SS contribute to *S. suis* virulence (serotypes 9 and 2, respectively) by promoting anti-phagocytic activity and the release of proinflammatory cytokines (Jiang et al., 2016; Auger et al., 2019). The agglutinin receptor was present in 86% of the commensal pathotype compared to 72% of the pathogenic pathotype while the T4SS components were widely distributed among the three pathotypes (86-96%). Our results indicated a lack of correlation between the presence of MGE-associated VAGs and pathotype. The presence of VAGs in a majority of the commensal pathotype provides further evidence that commensal strains may act as gene reservoirs (von Wintersdorff et al., 2016; Rossi et al., 2020). Although the 89K PAI (a *S. suis* MGE) was absent in all isolates in this study, the *in silico* detection of 89K PAI sequences (CH1/CH2 and CH5/CH6) and the presence of PAI VAGs suggests some genetic similarity between U.S. isolates and the virulent *S. suis* strain 05ZYH33 (Schmid et al., 2011).

Genes involved in the maintenance of MGEs were also determined from the MGE drafts (as determined by BLAST and prokka annotations) (Additional file 4.5). Replication proteins (helicases, primases, replication initiators), integrases/recombinase, and transposases are important for the formation and integration of MGEs (Gérard Guédon et al., 2017), and were annotated in 71-98% of our isolates with the last two

predominantly found in the pathogenic and possibly opportunistic pathotypes. Transcriptional regulators play a role in the expression of genes in and surrounding MGEs and were present in all three pathotypes (91-100%). Toxin-antitoxin and abortive infection proteins are responsible for the maintenance of resistance MGEs by promoting degradation of bacterial cells lacking the MGEs (Díaz-Orejas et al., 2017; Yang and Walsh, 2017). A putative zeta/pezT toxin (*SSU05_0936*) was predominantly identified in the pathogenic and possibly opportunistic pathotypes (70-72%). Many of the MGE drafts in our study contained genes encoding major carbohydrate transport systems (PTS transporters) (28%), a mobility protein (methyl-accepting chemotaxis protein) (72%), and metal resistance proteins (arsenic, calcium) (76-77%). Together these findings agree with publications stating MGEs carry a range of genes that contribute to the survival and adaptation of pathogenic bacteria to dynamic environments (Rankin et al., 2011; Durrant et al., 2020). Our results suggest MGE-mediated transfer of beneficial genes, including resistance genes and VAGs, is possible in U.S. *S. suis* isolates.

Comparative genome analysis of 208 *S. suis* isolates demonstrated a potential enrichment of virulence-associated genes in the accessory genome and elucidated a novel VAG genotyping scheme (*SSU_RS09155*+/*SSU_RS09525*+/*SSU_RS03100*+) for identifying pathogenic *S. suis* strains in the United States. We further determined the diversity of MGEs in the training set and determined that MGEs have the potential to gain/or spread resistance genes and putative VAGs between *S. suis* strains. Further research is needed *in vitro* to evaluate the contribution of the proposed VAGs to virulence.

4.6 Acknowledgements

The authors would like to thank Aaron Rendahl (University of Minnesota, St. Paul, MN) for providing statistical assistance.

4.7 Funding

The research project was partially funded by the Minnesota State Legislature through the Rapid Agricultural Response Fund, managed by the Minnesota Agricultural Experiment Station (project #00057268). April A. Estrada's stipend was funded by the USDA National Needs Fellowship (2016-38420-25288), University of Minnesota (UMN) MnDRIVE Global Food Ventures Fellowship, and UMN Informatics Institute Fellowship awards.

4.8 Figures

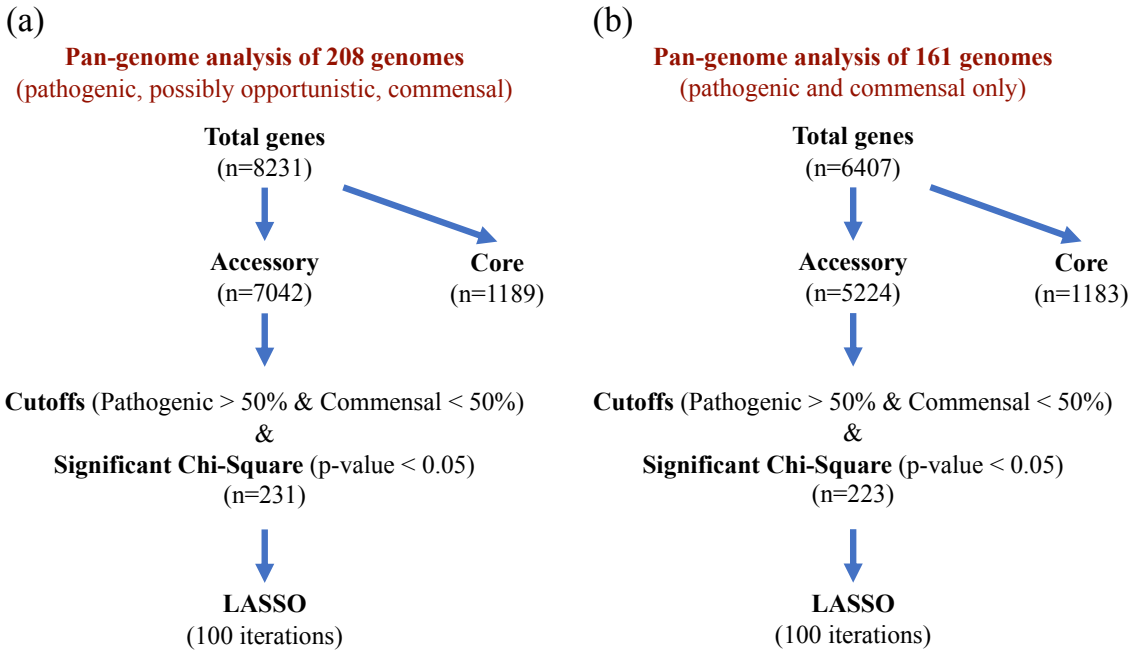

Figure 4.1 Illustration of the pan-genome approach for identifying candidate virulence-associated genes (VAGs). (a) The first data set analyzed consisted of all 208 isolates and **(b)** the second data set consisted of a subset of 161 isolates. Genes with a significant association (p-value < 0.05) with pathotype, and present in more than 50% (69/139) of isolates within the pathogenic pathotype and present in less than 50% (11/22) of isolates within the commensal pathotype, were analyzed using the LASSO shrinkage regression model. Genes identified in both LASSO analyses were selected as the 'best' predictors of pathotype.

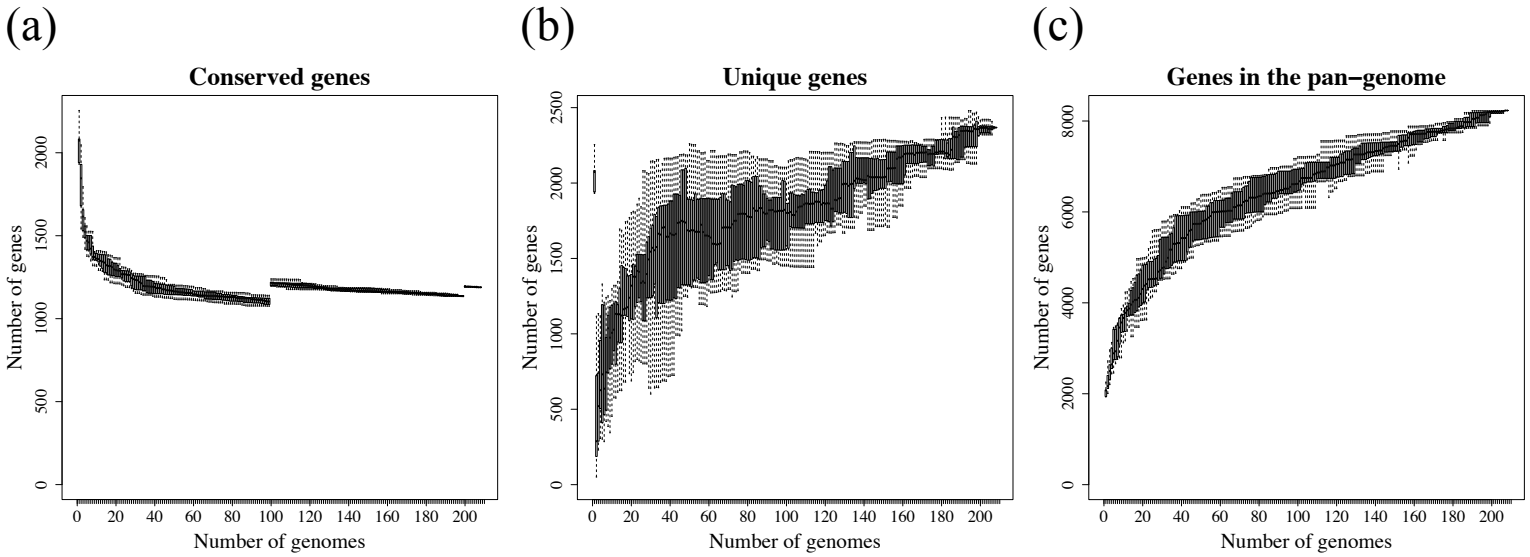

Figure 4.2 Summary statistics for the pan-genome analysis of 208 *S. suis* isolates. (a) Number of conserved genes, **(b)** unique genes, and **(c)** total number of genes plotted against the number of genomes included in the analysis.

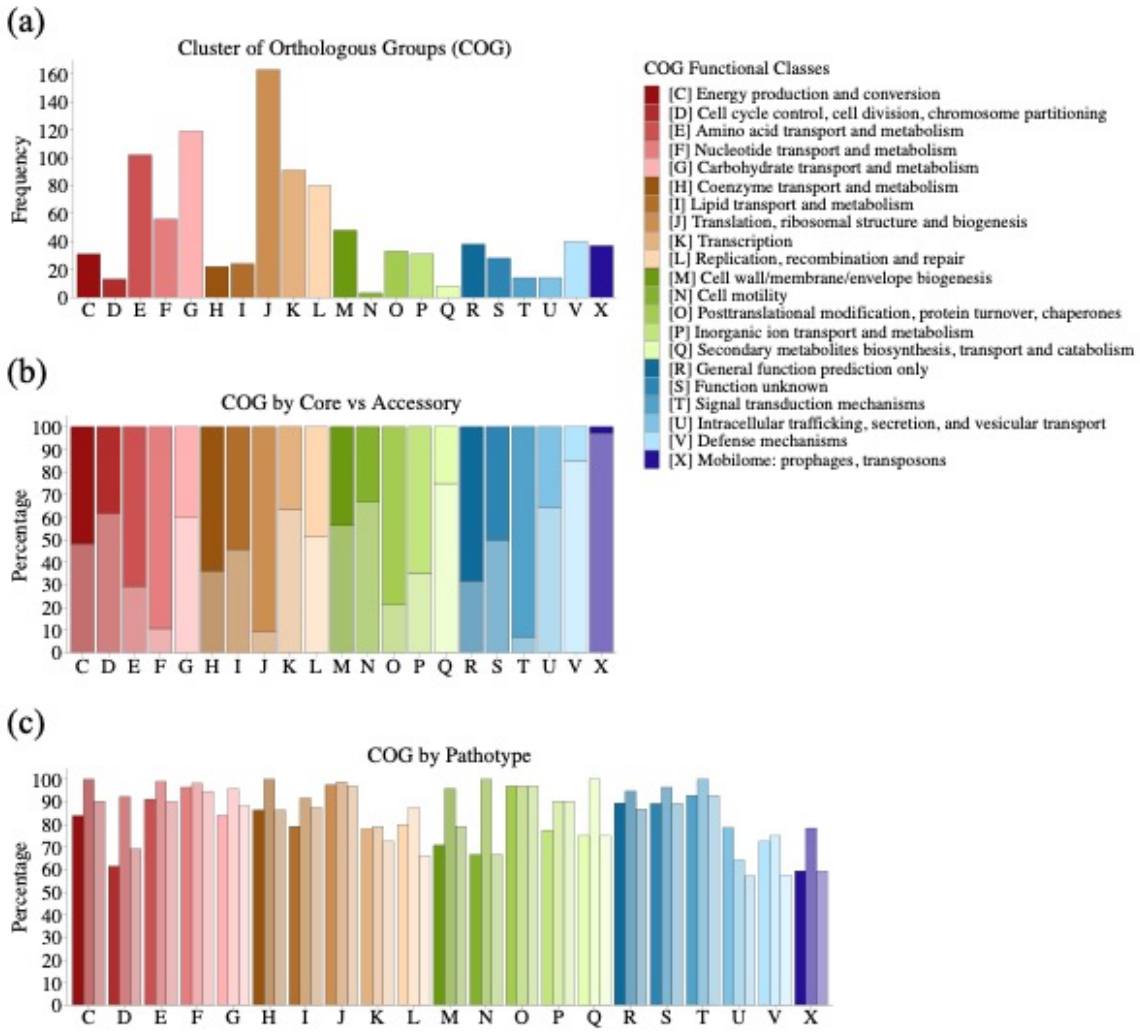

Figure 4.3 Predicted Cluster of Orthologous Groups (COG) functional classes. COG functional classes were determined for the 995 pan-gene clusters that met the criteria (psiblast, coverage ≥ 70%, percent identity ≥ 70%, and e-value ≤ 10-5). **(a)** The total frequencies of pan-gene clusters classified into each COG functional class (single-letter code) were plotted and color-coded. **(b)** The percentages of accessory and core pan-gene clusters classified into each COG functional class were plotted. **(c)** The percentages of pan-gene clusters (from each COG class) present in each pathotype (pathogenic, possibly opportunistic, commensal).

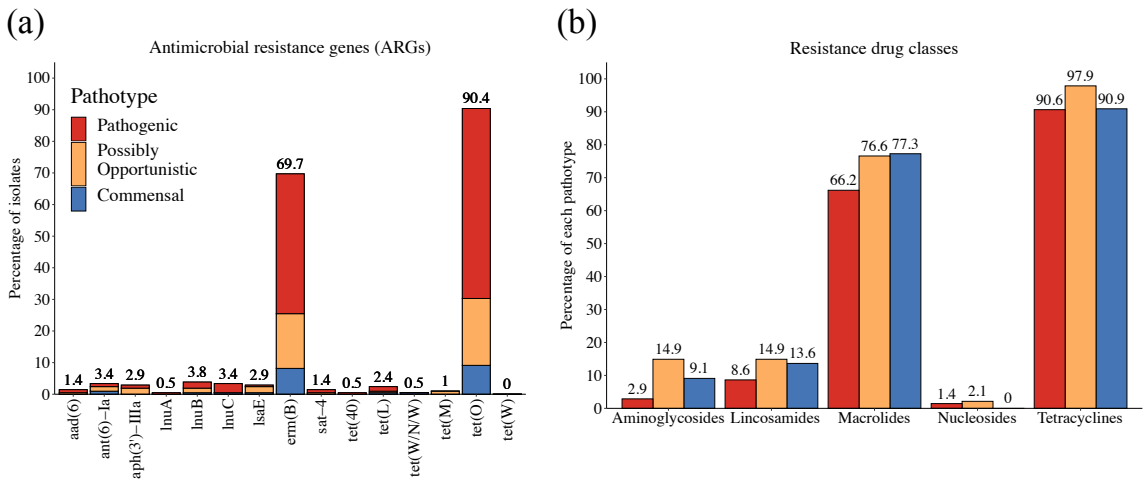

Figure 4.4 Antimicrobial resistance genes (ARGs) and drug classes determined in *S. suis* isolates (n=208). (a) Percentage of isolates possessing each of the 15 ARGs identified, conferring resistance to aminoglycosides, lincosamides/macrolides, nucleosides, and tetracyclines. ARGs were subdivided by pathotype (stacked bar sections; pathogenic, possibly opportunistic, and commensal). **(b)** Percentage of each pathotype possessing ARGs conferring resistance to each of the four drug classes.

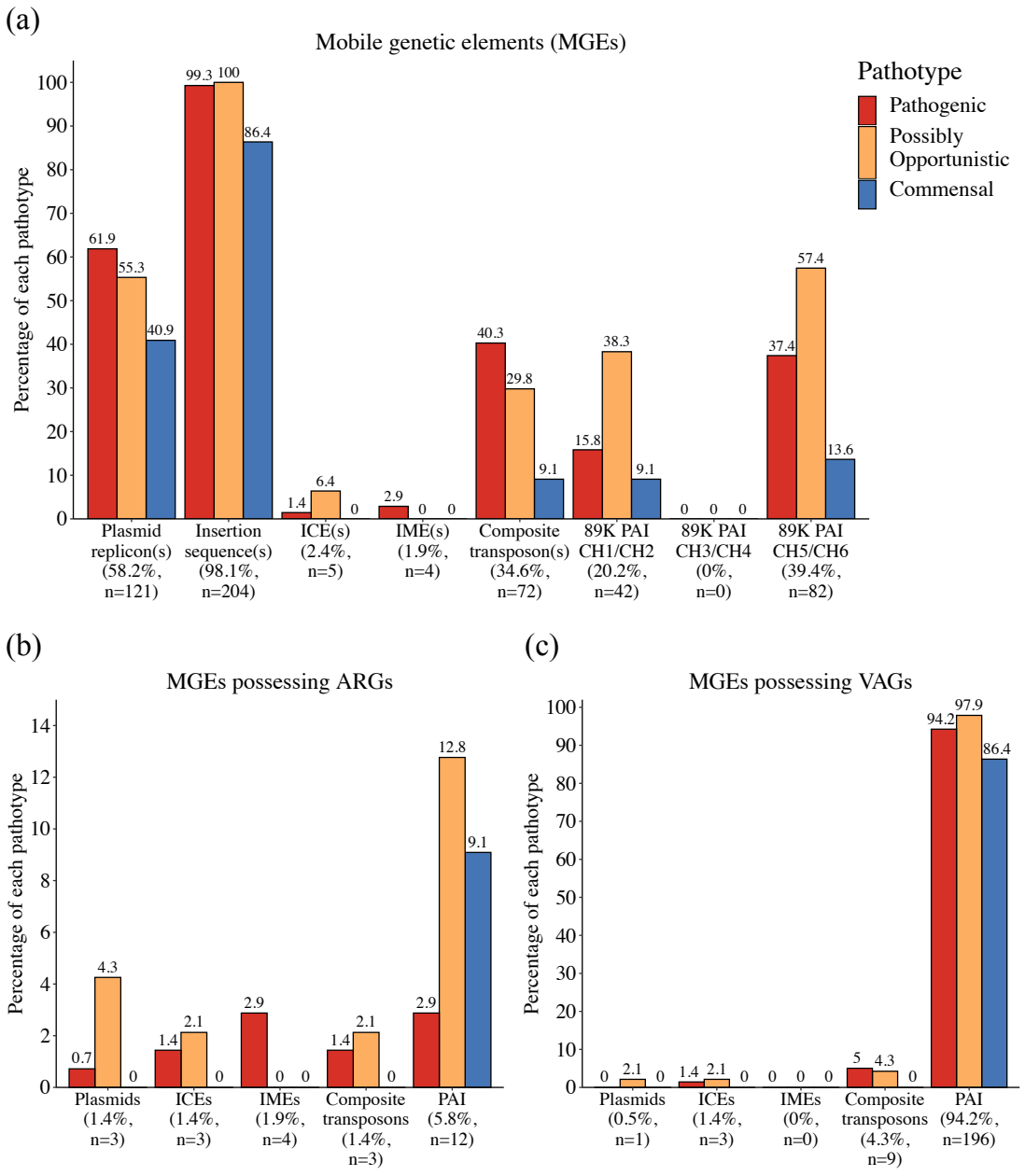

Figure 4.5 Mobile genetic elements carrying ARGs and VAGs. Percentage of each pathotype (pathogenic, possibly opportunistic, and commensal) possessing **(a)** various MGE types, **(b)** ARG-associated MGEs, and **(c)** VAG-associated MGEs. *S. suis* strain 05ZYH33 89K PAI sequences CH1/CH2 (5' flanking region), CH3/CH4 (716bp internal fragment), CH5/CH6 (3' flanking region). Total percentage and number of isolates possessing each type of MGE is listed under its respective name.

4.9 Tables

Table 4.1 Candidate VAGs identified by two LASSO analyses of S. *suis* genomes

Pan-gene group* in analysis of 208 genomes	**Pan-gene group* in analysis of 161 genomes**	***S. suis* strain reference**	**Size (AA)**	**Annotation**
group_766[a]	group_584[a]	SSU_RS09155 (P1/7) (WP_012028544.1)	219	hypothetical protein
group_790[a]	group_739[a]	SSU_RS01590 (P1/7) (WP_012774960.1)	175	membrane protein/ECF transporter S component
group_878[a]	group_693[a]	SSU_RS09525 (P1/7) (WP_012027987.1)	137	RNA-binding protein
group_1486[a]	group_1385[a]	SSU_RS03100 (P1/7) (WP_012775033.1)	104	hypothetical protein
group_1568[b]	group_1492	SSU_RS02325 (P1/7) (WP_004194478.1)	86	DNA repair protein
group_2061[b]	group_1532	SSU0541 (P1/7) (CAR45198.1)	116	transposase
group_483[b]	group_551	WP_002937844.1	116	hypothetical protein
group_1187	group_797[c]	SSU_RS01165 (P1/7) (WP_011921803.1)	343	ethanolamine utilization protein EutG
group_1475	group_1038[c]	SSU_RS03675 (P1/7) (WP_012775065.1)	150	hypothetical protein
group_1993	group_2992[c]	WP_012027348.1	76	hypothetical protein

[a] Pan-genes identified in both LASSO analyses that were selected as predictors of pathogenicity

[b] Candidate VAGs identified by LASSO analysis of 208 genomes

[c] Candidate VAGs identified by LASSO analysis of 161 genomes (only pathogenic and commensal pathotypes)

***** Group name is unique and determined by each ROARY analysis. Hence, group names are different for the same pan-gene cluster.

Table 4.2 LASSO results for the four candidate VAGs in the pathotypes of *S. suis* isolates (*n* = 208)

Candidate VAG(s)	No. containing the candidate VAG(s)	Pathogenic (*n* = 139)		Possibly Opportunistic (*n* = 47)		Commensal (*n* = 22)	
		No.	Proportion*	No.	Proportion*	No.	Proportion*
SSU_RS09155	178	137	0.770	34	0.191	7	0.039
SSU_RS01590	167	127	0.760	31	0.186	9	0.054
SSU_RS09525	189	138	0.730	42	0.222	9	0.048
SSU_RS03100	172	134	0.779	34	0.198	4	0.023
SSU_RS09155 SSU_RS09525 SSU_RS03100	168	134	0.798	31	0.185	3	0.018

* Positive isolates in the pathotype divided by the number of isolates containing the candidate VAG(s)

Table 4.3 Presence of novel candidate VAGs in virulent Eurasian *S. suis* strains

			% identity						
Strain	**Serotype**	**ST**	**SSU_RS09155**	**SSU_RS09525**	**SSU_RS03100**	**Source**	**Origin**	**Accession no.**	**Ref**
GZ1	2	1	96.1	98.8	100	human	China	CP000837	(Ye et al., 2009)
SC84	2	7	96.1	98.8	100	human	China	FM252031	(Holden et al., 2009)
P1/7	2	1	96.1	98.8	100	diseased pig	United Kingdom	AM946016	(Holden et al., 2009)
S735	2	1	96.1	98.8	100	diseased pig	The Netherlands	CP003736	(Boyle et al., 2012)
ZY05719	2	7	96.1	98.8	100	diseased pig	China	CP007497	(Zhang and Lu, 2007)
SC19	2	7	96.1	98.8	99.7	diseased pig	China	CP020863	(Li et al., 2010a)
05ZYH33	2	7	96.1	97.4	100	human	China	CP000407	(Chen et al., 2007)
10	2	1	96.1	98.8	100	healthy pig	The Netherlands	CP058742	(Vecht et al., 1992; Bunk et al., 2021)

GENERAL CONCLUSIONS

This dissertation studied the epidemiology and genomic characterization of *S. suis* and various methods for differentiating pathogenic strains. This was in an effort to address the lack of reports on *S. suis* diversity and pathogenicity in the United States. Our approach was modeled off traditional methods utilized in Europe, Asia, and Canada for characterizing virulent porcine and human clinical isolates.

In Chapter 2 we elucidated the heterogeneity of *S. suis* isolates originating from US swine herds by serotyping and MLST and defined a pathotype classification system for investigating links between subtypes (serotype and ST) and *S. suis* disease. U.S. isolates were highly diverse, representing 20 different serotypes and 58 different STs. Furthermore, subtype distributions were similar to that of Canada but differed from that of European and Asian countries. Investigation by pathotype classification revealed that only some subtypes were uniquely associated with diseased animals, supporting previous studies that found that despite the heterogeneity of *S. suis*, a small number of subtypes are likely responsible for a majority of clinical cases. Although unique pathogenic subtypes, such as SS1/2 and ST28, were prevalent in the United States, globally reported virulent strains, particularly SS2 and ST1, were also identified and should be monitored. In addition, a few subtypes were uniquely associated with healthy animals and a large number of novel STs were discovered in commensal and possibly opportunistic isolates, indicating commensal and opportunistic strains contribute to the high genetic diversity of *S. suis*.

In Chapter 3 we evaluated the distribution of previously published *S. suis* VAGs and their associations with pathotypes, subtypes, and epidemiological origin (swine production companies) of the isolates. We first demonstrated that a majority of the pathogenic pathotype lacked the three "classical" virulence markers of SS2 Eurasian strains (*epf*, *mrp*, and *sly*) and that U.S. pathogenic strains likely possess different VAGs. However, a few Eurasian-type strains were found in the United States, as illustrated in

Chapter 2, and were detected using the classical VAGs. Therefore, the *epf*, *mrp*, and *sly* genes remain valuable virulence markers for *S. suis*. General clustering of isolates by VAG profiles demonstrated that pathogenic strains possessed more VAGs than the commensals and that there were candidate genes capable of discriminating between pathotypes. We investigated candidate VAGs by statistical analyses and determined that the *ofs* and *srtF* genes better serve as predictors of pathogenicity than the presence of the classical VAGs. Thus, the incorporation of the *ofs* and *srtF* genes in diagnostic testing may improve the subtyping and detection of strains with increased virulence potential, including non-SS2 strains that are prevalent in U.S. swine herds. The comparison of *S. suis* isolates originating from different production companies depicted high diversity within and between production companies, and there was a lack of association between genotypes and production company. This suggests that strains of clinical significance may be widespread in the United States.

In Chapter 4 we investigated novel candidate virulence-associated genes and predicted antimicrobial resistance genes and mobile genetic elements of U.S. *S. suis* isolates. Functional annotation of the pan-genome provided further evidence that the *S. suis* accessory genome possesses genes with virulence-related functions and may be prone to horizontal gene transfer (mobilome). We identified four accessory pan-genes with a strong association with the pathogenic pathotype. Three of these genes, corresponding to the *S. suis* strain P1/7 genes *SSU_RS09155*, *SSU_RS09525*, and *SSU_RS03100*, were shown to be promising novel markers for U.S. pathogenic strains and potentially global strains. These markers, along with the classical and newly proposed markers described in Chapter 2, may serve as targets for diagnostic tests. In general, our approach helped identify genes and their predicted encoded proteins that may play a role in the pathogenesis of *S. suis* disease and may contribute to virulence in different serotype and ST strains, and thus should be investigated further *in vitro* and/or *in vivo*. High resistance to antibiotic classes widely used in the pig industry, but not typically used to treat *S. suis* infections, was predicted (tetracycline and erythromycin resistance genes). No beta-lactam resistance genes were predicted, suggesting a

continuing susceptibility to major treatment drugs in *S. suis*. One limitation, however, is we did not investigate gene mutations conferring antibiotic resistance. Overall, several isolates possessed MGEs and horizontal transfer of ARGs (and VAGs) is possible but did not appear to play a major role. MGE-linked resistance genes were predominantly present in commensal and possibly opportunistic isolates which are frequently exposed to therapeutic and nontherapeutic antibiotics.

In conclusion, this dissertation has contributed to a better understanding of *S. suis* in the United States. We described the distributions and prevalence of pathogenic subtypes that may be contributing to the persistence and severity of *S. suis* infections in swine herds. Previously published and novel VAGs proposed through our research may serve as targets for diagnostics or for experimental studies testing their roles in virulence or as vaccine candidates. Additionally, the WGS-based approach can be utilized for characterizing isolates from a particular farm or production system and aid in the monitoring and targeting of relevant strains for control and prevention strategies. Finally, *S. suis* disease is complex and multifactorial. The development and standardization of animal models for defining *S. suis* virulence and the international agreement on virulent strains are some recommendations that *S. suis* researchers should keep in mind. This could help reduce the variability in results produced by different research groups, therefore, improving the study of pathogenesis and virulence and the discovery of relevant virulence factors.

Finally, future research directions include the development of *in vitro* and/or *in vivo* tests for confirming the virulence of suspected pathogenic strains from the United States. As mentioned previously, the presence of virulence-associated factors does not guarantee virulence, thus prevalent pathogenic strains, such as serotype 1/2 ST28, should be evaluated alongside suspected commensal strains. Proposed VAGs from this dissertation may be further tested in a global collection of *S. suis* isolates including agreed-upon virulent and avirulent references. This would provide information on whether the proposed VAGs are specific to U.S. strains or may be used as global markers

of virulence. Characterization of proteins encoded by the novel VAGs (Chapter 4) and evaluation of their immunogenic properties may provide evidence on whether the novel virulence-associated factors can serve as vaccine candidates.

REFERENCES

Alcock, B.P., Raphenya, A.R., Lau, T.T.Y., Tsang, K.K., Bouchard, M., Edalatmand, A., Huynh, W., Nguyen, A.-L.V., Cheng, A.A., Liu, S., Min, S.Y., Miroshnichenko, A., Tran, H.-K., Werfalli, R.E., Nasir, J.A., Oloni, M., Speicher, D.J., Florescu, A., Singh, B., Faltyn, M., Hernandez-Koutoucheva, A., Sharma, A.N., Bordeleau, E., Pawlowski, A.C., Zubyk, H.L., Dooley, D., Griffiths, E., Maguire, F., Winsor, G.L., Beiko, R.G., Brinkman, F.S.L., Hsiao, W.W.L., Domselaar, G.V., McArthur, A.G., 2020. CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. Nucleic Acids Res 48, D517–D525. doi:10.1093/nar/gkz935

Allen, A.G., Bolitho, S., Lindsay, H., Khan, S., Bryant, C., Norton, P., Ward, P., Leigh, J., Morgan, J., Riches, H., Eastty, S., Maskell, D., 2001. Generation and characterization of a defined mutant of *Streptococcus suis* lacking suilysin. Infect. Immun. 69, 2732–2735. doi:10.1128/IAI.69.4.2732-2735.2001

Allen, J.P., Ozer, E.A., Minasov, G., Shuvalova, L., Kiryukhina, O., Satchell, K.J.F., Hauser, A.R., 2020. A comparative genomics approach identifies contact-dependent growth inhibition as a virulence determinant. Proc Natl Acad Sci U S A 117, 6811–6821. doi:10.1073/pnas.1919198117

Allgaier, A., Goethe, R., Wisselink, H.J., Smith, H.E., Valentin-Weigand, P., 2001. Relatedness of *Streptococcus suis* isolates of various serotypes and clinical backgrounds as evaluated by macrorestriction analysis and expression of potential virulence traits. J. Clin. Microbiol. 39, 445–453. doi:10.1128/JCM.39.2.445-453.2001

Amarasinghe, S.L., Su, S., Dong, X., Zappia, L., Ritchie, M.E., Gouil, Q., 2020. Opportunities and challenges in long-read sequencing data analysis. Genome Biol 21, 30. doi:10.1186/s13059-020-1935-5

Amass, S.F., SanMiguel, P., Clark, L.K., 1997. Demonstration of vertical transmission of *Streptococcus suis* in swine by genomic fingerprinting. J Clin Microbiol 35, 1595–1596. doi:10.1128/jcm.35.6.1595-1596.1997

Andam, C.P., Hanage, W.P., 2015. Mechanisms of genome evolution of *Streptococcus*. Infection, Genetics and Evolution 33, 334–342. doi:10.1016/j.meegid.2014.11.007

Anisimova, M., Bielawski, J., Dunn, K., Yang, Z., 2007. Phylogenomic analysis of natural selection pressure in *Streptococcus* genomes. BMC Evol Biol 7, 154. doi:10.1186/1471-2148-7-154

Aranda, J., Garrido, M.E., Fittipaldi, N., Cortés, P., Llagostera, M., Gottschalk, M., Barbé, J., 2009. Protective capacities of cell surface-associated proteins of *Streptococcus suis* mutants deficient in divalent cation-uptake regulators. Microbiology 155, 1580–1587. doi:10.1099/mic.0.026278-0

Arenas, J., Bossers-de Vries, R., Harders-Westerveen, J., Buys, H., Ruuls-van Stalle, L.M.F., Stockhofe-Zurwieden, N., Zaccaria, E., Tommassen, J., Wells, J.M., Smith, H.E., de Greeff, A., 2019. *In vivo* transcriptomes of *Streptococcus suis*

reveal genes required for niche-specific adaptation and pathogenesis. Virulence 10, 334–351. doi:10.1080/21505594.2019.1599669

Arends, J.P., Hartwig, N., Rudolphy, M., Zanen, H.C., 1984. Carrier rate of *Streptococcus suis* capsular type 2 in palatine tonsils of slaughtered pigs. J. Clin. Microbiol. 20, 945–947. doi:10.1128/jcm.20.5.945-947.1984

Arends, J.P., Zanen, H.C., 1988. Meningitis caused by *Streptococcus suis* in humans. Clinical Infectious Diseases 10, 131–137. doi:10.1093/clinids/10.1.131

Arndt, E.R., Farzan, A., MacInnes, J.I., Friendship, R.M., 2019. Antimicrobial resistance of *Streptococcus suis* isolates recovered from clinically ill nursery pigs and from healthy pigs at different stages of production. Can Vet J 60, 519–522.

Athey, T.B.T., Auger, J.-P., Teatero, S., Dumesnil, A., Takamatsu, D., Wasserscheid, J., Dewar, K., Gottschalk, M., Fittipaldi, N., 2015. Complex population structure and virulence differences among serotype 2 *Streptococcus suis* strains belonging to sequence type 28. PLoS ONE 10, e0137760. doi:10.1371/journal.pone.0137760

Athey, T.B.T., Teatero, S., Lacouture, S., Takamatsu, D., Gottschalk, M., Fittipaldi, N., 2016. Determining *Streptococcus suis* serotype from short-read whole-genome sequencing data. BMC Microbiology 16, 162–169. doi:10.1186/s12866-016-0782-8

Auger, J.-P., Boa, A.-C., Segura, M., Gottschalk, M., 2019. Antigen I/II participates in the interactions of *Streptococcus suis* serotype 9 with phagocytes and the development of systemic disease. Front Cell Infect Microbiol 9, 124. doi:10.3389/fcimb.2019.00124

Auger, J.-P., Fittipaldi, N., Benoit-Biancamano, M.-O., Segura, M., Gottschalk, M., 2016a. Virulence studies of different sequence types and geographical origins of *Streptococcus suis* serotype 2 in a mouse model of infection. Pathogens 5, 48–71. doi:10.3390/pathogens5030048

Auger, J.-P., Fittipaldi, N., Benoit-Biancamano, M.-O., Segura, M., Gottschalk, M., 2016b. Virulence Studies of Different Sequence Types and Geographical Origins of *Streptococcus suis* Serotype 2 in a Mouse Model of Infection. Pathogens 5. doi:10.3390/pathogens5030048

Baums, C.G., Brüggemann, C., Kock, C., Beineke, A., Waldmann, K.-H., Valentin-Weigand, P., 2010. Immunogenicity of an autogenous *Streptococcus suis* bacterin in preparturient sows and their piglets in relation to protection after weaning. Clin. Vaccine Immunol. 17, 1589–1597. doi:10.1128/CVI.00159-10

Baums, C.G., Kaim, U., Fulde, M., Ramachandran, G., Goethe, R., Valentin-Weigand, P., 2006. Identification of a novel virulence determinant with serum opacification activity in *Streptococcus suis*. Infect Immun 74, 6154–6162. doi:10.1128/IAI.00359-06

Baums, C.G., Kock, C., Beineke, A., Bennecke, K., Goethe, R., Schröder, C., Waldmann, K.-H., Valentin-Weigand, P., 2009. *Streptococcus suis* bacterin and subunit vaccine immunogenicities and protective efficacies against serotypes 2 and 9. Clin Vaccine Immunol 16, 200–208. doi:10.1128/CVI.00371-08

Baums, C.G., Valentin-Weigand, P., 2009. Surface-associated and secreted factors of *Streptococcus suis* in epidemiology, pathogenesis and vaccine development. Anim Health Res Rev 10, 65–83. doi:10.1017/S146625230999003X

Beaudoin, M., Harel, J., Higgins, R., Gottschalk, M., Frenette, M., MacInnes, J.I., 1992. Molecular analysis of isolates of *Streptococcus suis* capsular type 2 by restriction-endonuclease-digested DNA separated on SDS-PAGE and by hybridization with an rDNA probe. J. Gen. Microbiol. 138, 2639–2645. doi:10.1099/00221287-138-12-2639

Beineke, A., Bennecke, K., Neis, C., Schröder, C., Waldmann, K.-H., Baumgärtner, W., Valentin-Weigand, P., Baums, C.G., 2008. Comparative evaluation of virulence and pathology of *Streptococcus suis* serotypes 2 and 9 in experimentally infected growers. Veterinary Microbiology 128, 423–430. doi:10.1016/j.vetmic.2007.10.028

Benga, L., Goethe, R., Rohde, M., Valentin-Weigand, P., 2004. Non-encapsulated strains reveal novel insights in invasion and survival of *Streptococcus suis* in epithelial cells. Cellular Microbiology 6, 867–881. doi:10.1111/j.1462-5822.2004.00409.x

Berthelot-Hérault, F., Gottschalk, M., Labbé, A., Cariolet, R., Kobisch, M., 2001. Experimental airborne transmission of *Streptococcus suis* capsular type 2 in pigs. Veterinary Microbiology 82, 69–80. doi:10.1016/S0378-1135(01)00376-5

Berthelot-Hérault, F., Marois, C., Gottschalk, M., Kobisch, M., 2002. Genetic diversity of *Streptococcus suis* strains isolated from pigs and humans as revealed by pulsed-field gel electrophoresis. J. Clin. Microbiol. 40, 615–619. doi:10.1128/jcm.40.2.615-6192002

Blakely, G.W., 2015. Mechanisms of horizontal gene transfer and DNA recombination, in: Molecular Medical Microbiology. Elsevier, pp. 291–302. doi:10.1016/B978-0-12-397169-2.00015-9

Bodenhofer, U., Bonatesta, E., Horejs-Kainrath, C., Hochreiter, S., 2015. msa: an R package for multiple sequence alignment. Bioinformatics 31, 3997–3999. doi:10.1093/bioinformatics/btv494

Boetner, A.G., Binder, M., Bille-Hansen, V., 1987. *Streptococcus suis* infections in Danish pigs and experimental infection with *Streptococcus suis* serotype 7. Acta Pathologica Microbiologica Scandinavica Series B: Microbiology 95B, 233–239. doi:10.1111/j.1699-0463.1987.tb03118.x

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120. doi:10.1093/bioinformatics/btu170

Bonifait, L., Gottschalk, M., Grenier, D., 2010. Cell surface characteristics of nontypeable isolates of *Streptococcus suis*. FEMS Microbiology Letters 311, 160–166. doi:10.1111/j.1574-6968.2010.02086.x

Bonifait, L., Grenier, D., 2011. The SspA subtilisin-like protease of *Streptococcus suis* triggers a pro-inflammatory response in macrophages through a non-proteolytic mechanism. BMC Microbiol. 11, 47. doi:10.1186/1471-2180-11-47

Bonifait, L., Veillette, M., Létourneau, V., Grenier, D., Duchaine, C., 2014. Detection of *Streptococcus suis* in bioaerosols of swine confinement buildings. Appl. Environ. Microbiol. 80, 3296–3304. doi:10.1128/AEM.04167-13

Bousis, S., Setyawati, I., Diamanti, E., Slotboom, D.J., Hirsch, A.K.H., 2019. Energy-coupling factor transporters as novel antimicrobial targets. Adv. Therap. 2, 3296–3304. doi:10.1002/adtp.201800066

Boyle, B., Vaillancourt, K., Bonifait, L., Charette, S.J., Gottschalk, M., Grenier, D., 2012. Genome sequence of the swine pathogen *Streptococcus suis* serotype 2 strain S735. J Bacteriol 194, 6343–6344. doi:10.1128/JB.01559-12

Brambila-Tapia, A.J.L., Armenta-Medina, D., Rivera-Gomez, N., Perez-Rueda, E., 2014. Main functions and taxonomic distribution of virulence genes in *Brucella melitensis* 16 M. PLoS One 9, e100349. doi:10.1371/journal.pone.0100349

Brassard, J., Gottschalk, M., Quessy, S., 2004. Cloning and purification of the *Streptococcus suis* serotype 2 glyceraldehyde-3-phosphate dehydrogenase and its involvement as an adhesin. Vet. Microbiol. 102, 87–94. doi:10.1016/j.vetmic.2004.05.008

Brisebois, L.M., Charlebois, R., Higgins, R., Nadeau, M., 1990. Prevalence of *Streptococcus suis* in four to eight week old clinically healthy piglets. Can. J. Vet. Res. 54, 174–177.

Brockmeier, S.L., Loving, C.L., Nicholson, T.L., Wang, J., Peters, S.E., Weinert, L., Chaudhuri, R., Seilly, D.J., Langford, P.R., Rycroft, A., Wren, B.W., Maskell, D.J., Tucker, A.W., 2018. Use of proteins identified through a functional genomic screen to develop a protein subunit vaccine that provides significant protection against virulent *Streptococcus suis* in pigs. Infect Immun 86, e00559-17. doi:10.1128/IAI.00559-17

Bunk, B., Jakóbczak, B., Florian, V., Dittmar, D., Mäder, U., Jarek, M., Häußler, S., Baums, C.G., Völker, U., Michalik, S., 2021. Complete genome sequences of *Streptococcus suis* pig-pathogenic strains 10, 13-00283-02, and 16085/3B. Microbiol Resour Announc 10, e01137-20. doi:10.1128/MRA.01137-20

Busque, P., Higgins, R., Caya, F., Quessy, S., 1997. Immunization of pigs against *Streptococcus suis* serotype 2 infection using a live avirulent strain. Can J Vet Res 61, 275–279.

Büttner, N., Beineke, A., de Buhr, N., Lilienthal, S., Merkel, J., Waldmann, K.-H., Valentin-Weigand, P., Baums, C.G., 2012. *Streptococcus suis* serotype 9 bacterin immunogenicity and protective efficacy. Veterinary Immunology and Immunopathology 146, 191–200. doi:10.1016/j.vetimm.2012.03.012

Calderaro, F.F., Moreno, L.Z., Doto, D.S., Matajira, C.E.C., Gomes, V.T.M., Ferreira, T.S.P., Mesquita, R.E., Moreno, A.M., 2016. Characterization of *Streptococcus suis* through serotyping, SE-AFLP and virulence profile. Pesq. Vet. Bras. 36, 701–704. doi:10.1590/S0100-736X2016000800004

Callejo, R., Zheng, H., Du, P., Prieto, M., Xu, J., Zielinski, G., Auger, J.-P., Gottschalk, M., 2016. *Streptococcus suis* serotype 2 strains isolated in Argentina (South America) are different from those recovered in North America and present a higher risk for humans. JMM Case Reports 3, e005066. doi:10.1099/jmmcr.0.005066

Calzas, C., Taillardet, M., Fourati, I., Roy, D., Gottschalk, M., Soudeyns, H., Defrance, T., Segura, M., 2017. Evaluation of the immunomodulatory properties of

Streptococcus suis and group B *Streptococcus* capsular polysaccharides on the humoral response. Pathogens 6, 16. doi:10.3390/pathogens6020016
Cantin, M., Harel, J., Higgins, R., Gottschalk, M., 1992. Antimicrobial resistance patterns and plasmid profiles of *Streptococcus suis* isolates. J Vet Diagn Invest 4, 170–174. doi:10.1177/104063879200400209
Cao, P., Guo, D., Liu, J., Jiang, Q., Xu, Z., Qu, L., 2017. Genome-wide analyses reveal genes subject to positive selection in *Pasteurella multocida*. Front. Microbiol. 8, 961. doi:10.3389/fmicb.2017.00961
Carattoli, A., Zankari, E., García-Fernández, A., Voldby Larsen, M., Lund, O., Villa, L., Møller Aarestrup, F., Hasman, H., 2014. In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. Antimicrob Agents Chemother 58, 3895–3903. doi:10.1128/AAC.02412-14
Carriço, J.A., Sabat, A.J., Friedrich, A.W., Ramirez, M., on behalf of the ESCMID Study Group, C., 2013. Bioinformatics in bacterial molecular epidemiology and public health: databases, tools and the next-generation sequencing revolution. Eurosurveillance 18, 20382–20390. doi:10.2807/ese.18.04.20382-en
Chabot-Roy, G., Willson, P., Segura, M., Lacouture, S., Gottschalk, M., 2006. Phagocytosis and killing of *Streptococcus suis* by porcine neutrophils. Microb Pathog. 41, 21–32. doi:10.1016/j.micpath.2006.04.001
Charif, D., Lobry, J.R., 2007. SeqinR 1.0-2: a contributed package to the R project for statistical computing devoted to biological sequences retrieval and analysis, in: Bastolla, U., Porto, M., Roman, H.E., Vendruscolo, M. (Eds.), Structural Approaches to Sequence Evolution: Molecules, Networks, Populations, Biological and Medical Physics, Biomedical Engineering. Springer Verlag, New York, pp. 207–232.
Charland, N., Nizet, V., Rubens, C.E., Kim, K.S., Lacouture, S., Gottschalk, M., 2000. *Streptococcus suis* serotype 2 interactions with human brain microvascular endothelial cells. Infect. Immun. 68, 637–643.
Chatellier, S., Gottschalk, M., Higgins, R., Brousseau, R., Harel, J., 1999. Relatedness of *Streptococcus suis* serotype 2 isolates from different geographic origins as evaluated by molecular fingerprinting and phenotyping. Journal of Clinical Microbiology 37, 362–366. doi:10.1128/JCM.37.2.362-366.1999
Chatellier, S., Harel, J., Zhang, Y., Gottschalk, M., Higgins, R., Devriese, L.A., Brousseau, R., 1998. Phylogenetic diversity of *Streptococcus suis* strains of various serotypes as revealed by 16S rRNA gene sequence comparison. International Journal of Systematic Bacteriology 48, 581–589. doi:10.1099/00207713-48-2-581
Chen, C., Tang, J., Dong, W., Wang, C., Feng, Y., Wang, Jing, Zheng, F., Pan, X., Liu, D., Li, M., Song, Y., Zhu, X., Sun, H., Feng, T., Guo, Z., Ju, A., Ge, J., Dong, Y., Sun, W., Jiang, Y., Wang, Jun, Yan, J., Yang, H., Wang, X., Gao, G.F., Yang, R., Wang, Jian, Yu, J., 2007. A glimpse of streptococcal toxic shock syndrome from comparative genomics of *S. suis* 2 Chinese isolates. PLoS One 2, e315. doi:10.1371/journal.pone.0000315

Chen, C., Zhang, W., Zheng, H., Lan, R., Wang, H., Du, P., Bai, X., Ji, S., Meng, Q., Jin, D., Liu, K., Jing, H., Ye, C., Gao, G.F., Wang, L., Gottschalk, M., Xu, J., 2013. Minimum core genome sequence typing of bacterial pathogens: a unified approach for clinical and public health microbiology. Journal of Clinical Microbiology 51, 2582–2591. doi:10.1128/JCM.00535-13

Chen, S.L., Hung, C.-S., Xu, J., Reigstad, C.S., Magrini, V., Sabo, A., Blasiar, D., Bieri, T., Meyer, R.R., Ozersky, P., Armstrong, J.R., Fulton, R.S., Latreille, J.P., Spieth, J., Hooton, T.M., Mardis, E.R., Hultgren, S.J., Gordon, J.I., 2006. Identification of genes subject to positive selection in uropathogenic strains of *Escherichia coli*: a comparative genomics approach. Proceedings of the National Academy of Sciences 103, 5977–5982. doi:10.1073/pnas.0600938103

Chen, W., Simpson, J., Levesque, C.A., 2018. RAM: R for amplicon-sequencing-based microbial-ecology. Agriculture and Agri-Food Canada / Agriculture et Agroalimentaire Canada.

Clifton-Hadley, F.A., 1984. Studies of *Streptococcus suis* type 2 infection in pigs. Vet. Res. Commun. 8, 217–227.

Clifton-Hadley, F.A., Alexander, T.J., 1980. The carrier site and carrier rate of *Streptococcus suis* type II in pigs. Vet. Rec. 107, 40–41.

Cloutier, G., D'Allaire, S., Martinez, G., Surprenant, C., Lacouture, S., Gottschalk, M., 2003. Epidemiology of *Streptococcus suis* serotype 5 infection in a pig herd with and without clinical disease. Veterinary Microbiology 97, 135–151. doi:10.1016/j.vetmic.2003.09.018

Corsaut, L., Martelet, L., Goyette-Desjardins, G., Beauchamp, G., Denicourt, M., Gottschalk, M., Segura, M., 2021. Immunogenicity study of a *Streptococcus suis* autogenous vaccine in preparturient sows and evaluation of passive maternal immunity in piglets. BMC Vet Res 17, 72. doi:10.1186/s12917-021-02774-4

Corsaut, L., Misener, M., Canning, P., Beauchamp, G., Gottschalk, M., Segura, M., 2020. Field study on the immunological response and protective effect of a licensed autogenous vaccine to control *Streptococcus suis* infections in post-weaned piglets. Vaccines 8, 384. doi:10.3390/vaccines8030384

D'Auria, G., Jiménez-Hernández, N., Peris-Bondia, F., Moya, A., Latorre, A., 2010. *Legionella pneumophila* pangenome reveals strain-specific virulence factors. BMC Genomics 11, 181. doi:10.1186/1471-2164-11-181

Day, A., 2012. heatmap.plus: Heatmap with more sensible behavior.

de Greeff, A., Wisselink, H.J., de Bree, F.M., Schultsz, C., Baums, C.G., Thi, H.N., Stockhofe-Zurwieden, N., Smith, H.E., 2011. Genetic diversity of *Streptococcus suis* isolates as determined by comparative genome hybridization. BMC Microbiol. 11, 161–175. doi:10.1186/1471-2180-11-161

Dee, S.A., Carlson, A.R., Winkelman, N.L., Corey, M.M., 1993. Effect of management practices on the *Streptococcus suis* carrier rate in nursery swine. J Am Vet Med Assoc 203, 295–299.

Dekker, C.N.T., Bouma, A., Daemen, A.J.J.M., van Leengoed, L.A.M.G., Jonker, F.H., Wagenaar, J.A., Stegeman, J.A., 2012. Homologous whole bacterin vaccination is

not able to reduce *Streptococcus suis* serotype 9 strain 7997 transmission among pigs or colonization. Vaccine 30, 1379–1387. doi:10.1016/j.vaccine.2011.12.035
Dekker, N., Bouma, A., Daemen, I., Klinkenberg, D., van Leengoed, L., Wagenaar, J.A., Stegeman, A., 2013. Effect of spatial separation of pigs on spread of *Streptococcus suis* serotype 9. PLoS ONE 8, e61339. doi:10.1371/journal.pone.0061339
Denich, L.C., Farzan, A., Friendship, R., Arndt, E., Gottschalk, M., Poljak, Z., 2020. A case-control study to investigate the serotypes of *S. suis* isolates by multiplex PCR in nursery pigs in Ontario, Canada. Pathogens 9. doi:10.3390/pathogens9010044
Di Martino, P., 2018. Bacterial adherence: much more than a bond. AIMS Microbiology 4, 563–566. doi:10.3934/microbiol.2018.3.563
Díaz-Orejas, R., Espinosa, M., Yeo, C.C., 2017. The importance of the expendable: toxin–antitoxin genes in plasmids and chromosomes. Front. Microbiol. 8, 1479. doi:10.3389/fmicb.2017.01479
Dominguez-Punaro, M.C., Segura, M., Plante, M.-M., Lacouture, S., Rivest, S., Gottschalk, M., 2007. *Streptococcus suis* serotype 2, an important swine and human pathogen, induces strong systemic and cerebral inflammatory responses in a mouse model of infection. The Journal of Immunology 179, 1842–1854. doi:10.4049/jimmunol.179.3.1842
Dong, W., Ma, J., Zhu, Y., Zhu, J., Yuan, L., Wang, Y., Xu, J., Pan, Z., Wu, Z., Zhang, W., Lu, C., Yao, H., 2015. Virulence genotyping and population analysis of *Streptococcus suis* serotype 2 isolates from China. Infection, Genetics and Evolution 36, 483–489. doi:10.1016/j.meegid.2015.08.021
Dong, W., zhu, Y., Ma, Y., Ma, J., Zhang, Y., Yuan, L., Pan, Z., Wu, Z., Yao, H., 2017. Multilocus sequence typing and virulence genotyping of Streptococcus suis serotype 9 isolates revealed high genetic and virulence diversity. FEMS Microbiology Letters 364. doi:10.1093/femsle/fnx192
Doto, D.S., Moreno, L.Z., Calderaro, F.F., Matajira, C.E.C., de Moura Gomes, V.T., Ferreira, T.S.P., Mesquita, R.E., Timenetsky, J., Gottschalk, M., Moreno, A.M., 2016. Genetic diversity of *Streptococcus suis* serotype 2 isolated from pigs in Brazil. Can. J. Vet. Res. 80, 106–111.
Du, H., Huang, W., Xie, H., Ye, C., Jing, H., Ren, Z., Xu, J., 2013. The genetically modified suilysin, rSLYP353L, provides a candidate vaccine that suppresses proinflammatory response and reduces fatality following infection with *Streptococcus suis*. Vaccine 31, 4209–4215. doi:10.1016/j.vaccine.2013.07.004
Du, P., Zheng, H., Zhou, J., Lan, R., Ye, C., Jing, H., Jin, D., Cui, Z., Bai, X., Liang, J., Liu, J., Xu, L., Zhang, W., Chen, C., Xu, J., 2017. Detection of multiple parallel transmission outbreak of *Streptococcus suis* human infection by use of genome epidemiology, China, 2005. Emerging Infect. Dis. 23, 204–211. doi:10.3201/eid2302.160297
Dumesnil, A., Martelet, L., Grenier, D., Auger, J.-P., Harel, J., Nadeau, E., Gottschalk, M., 2019. Enolase and dipeptidyl peptidase IV protein sub-unit vaccines are not

protective against a lethal *Streptococcus suis* serotype 2 challenge in a mouse model of infection. BMC Vet Res 15, 448. doi:10.1186/s12917-019-2196-y

Durrant, M.G., Li, M.M., Siranosian, B.A., Montgomery, S.B., Bhatt, A.S., 2020. A bioinformatic analysis of integrative mobile genetic elements highlights their role in bacterial adaptation. Cell Host & Microbe 27, 140-153.e9. doi:10.1016/j.chom.2019.10.022

Dutkiewicz, J., Sroka, J., Zając, V., Wasiński, B., Cisak, E., Sawczyn, A., Kloc, A., Wójcik-Fatla, A., 2017. *Streptococcus suis*: a re-emerging pathogen associated with occupational exposure to pigs or pork products. Part I - Epidemiology. Ann Agric Environ Med 24, 683–695. doi:10.26444/aaem/79813

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797. doi:10.1093/nar/gkh340

Edwards, D.J., Holt, K.E., 2013. Beginner's guide to comparative bacterial genome analysis using next-generation sequence data. Microbial Informatics and Experimentation 3, 2–10. doi:10.1186/2042-5783-3-2

Eitinger, T., Rodionov, D.A., Grote, M., Schneider, E., 2011. Canonical and ECF-type ATP-binding cassette importers in prokaryotes: diversity in modular organization and cellular functions. FEMS Microbiol Rev 35, 3–67. doi:10.1111/j.1574-6976.2010.00230.x

Elliott, S.D., Clifton-Hadley, F., Tai, J., 1980. Streptococcal infection in young pigs. V. An immunogenic polysaccharide from *Streptococcus suis* type 2 with particular reference to vaccination against streptococcal meningitis in pigs. J. Hyg. 85, 275–285. doi:10.1017/S0022172400063312

Engelhardt, B., Sorokin, L., 2009. The blood–brain and the blood–cerebrospinal fluid barriers: function and dysfunction. Semin Immunopathol 31, 497–511. doi:10.1007/s00281-009-0177-0

Esgleas, M., Dominguez-Punaro, M. de la C., Li, Y., Harel, J., Dubreuil, J.D., Gottschalk, M., 2009. Immunization with SsEno fails to protect mice against challenge with *Streptococcus suis* serotype 2. FEMS Microbiology Letters 294, 82–88. doi:10.1111/j.1574-6968.2009.01551.x

Esgleas, M., Lacouture, S., Gottschalk, M., 2005. *Streptococcus suis* serotype 2 binding to extracellular matrix proteins. FEMS Microbiology Letters 244, 33–40. doi:10.1016/j.femsle.2005.01.017

Esgleas, M., Li, Y., Hancock, M.A., Harel, J., Dubreuil, J.D., Gottschalk, M., 2008. Isolation and characterization of alpha-enolase, a novel fibronectin-binding protein from *Streptococcus suis*. Microbiology (Reading, Engl.) 154, 2668–2679. doi:10.1099/mic.0.2008/017145-0

Estrada, A.A., Gottschalk, M., Rendahl, A., Rossow, S., Marshall-Lund, L., Marthaler, D.G., Gebhart, C.J., 2021. Proposed virulence-associated genes of *Streptococcus suis* isolates from the United States serve as predictors of pathogenicity. Porc Health Manag 7, 22. doi:10.1186/s40813-021-00201-6

Estrada, A.A., Gottschalk, M., Rossow, S., Rendahl, A., Gebhart, C., Marthaler, D.G., 2019. Serotype and genotype (Multilocus Sequence Type) of *Streptococcus suis*

isolates from the United States serve as predictors of pathotype. J. Clin. Microbiol. 57, e00377-19. doi:10.1128/JCM.00377-19
Facklam, R., 2002. What happened to the streptococci: overview of taxonomic and nomenclature changes. Clin Microbiol Rev 15, 613–630. doi:10.1128/CMR.15.4.613-630.2002
Faulds-Pain, A., Shaw, H.A., Terra, V.S., Kellner, S., Brockmeier, S.L., Wren, B.W., 2019. The *Streptococcos suis* sortases SrtB and SrtF are essential for disease in pigs. Microbiology (Reading) 165, 163–173. doi:10.1099/mic.0.000752
Feil, E.J., Li, B.C., Aanensen, D.M., Hanage, W.P., Spratt, B.G., 2004. eBURST: Inferring Patterns of Evolutionary Descent among Clusters of Related Bacterial Genotypes from Multilocus Sequence Typing Data. Journal of Bacteriology 186, 1518–1530. doi:10.1128/JB.186.5.1518-1530.2004
Feng, L., Niu, X., Mei, W., Li, W., Liu, Y., Willias, S.P., Yuan, C., Bei, W., Wang, X., Li, J., 2018. Immunogenicity and protective capacity of EF-Tu and FtsZ of *Streptococcus suis* serotype 2 against lethal infection. Vaccine 36, 2581–2588. doi:10.1016/j.vaccine.2018.03.079
Feng, Y., Pan, X., Sun, W., Wang, C., Zhang, H., Li, X., Ma, Y., Shao, Z., Ge, J., Zheng, F., Gao, G.F., Tang, J., 2009. *Streptococcus suis* enolase functions as a protective antigen displayed on the bacterial cell surface. J INFECT DIS 200, 1583–1592. doi:10.1086/644602
Feng, Y., Zhang, H., Wu, Z., Wang, S., Cao, M., Hu, D., Wang, C., 2014. *Streptococcus suis* infection: An emerging/reemerging challenge of bacterial infectious diseases? Virulence 5, 477–497. doi:10.4161/viru.28595
Ferrando, M.L., de Greeff, A., van Rooijen, W.J.M., Stockhofe-Zurwieden, N., Nielsen, J., Wichgers Schreur, P.J., Pannekoek, Y., Heuvelink, A., van der Ende, A., Smith, H., Schultsz, C., 2015. Host-pathogen interaction at the intestinal mucosa correlates with zoonotic potential of *Streptococcus suis*. The Journal of Infectious Diseases 212, 95–105. doi:10.1093/infdis/jiu813
Fittipaldi, N., Fuller, T.E., Teel, J.F., Wilson, T.L., Wolfram, T.J., Lowery, D.E., Gottschalk, M., 2009. Serotype distribution and production of muramidase-released protein, extracellular factor and suilysin by field strains of *Streptococcus suis* isolated in the United States. Vet. Microbiol. 139, 310–317. doi:10.1016/j.vetmic.2009.06.024
Fittipaldi, N., Harel, J., D'Amours, B., Lacouture, S., Kobisch, M., Gottschalk, M., 2007. Potential use of an unencapsulated and aromatic amino acid-auxotrophic *Streptococcus suis* mutant as a live attenuated vaccine in swine. Vaccine 25, 3524–3535. doi:10.1016/j.vaccine.2007.01.084
Fittipaldi, N., Segura, M., Grenier, D., Gottschalk, M., 2012. Virulence factors involved in the pathogenesis of the infection caused by the swine pathogen and zoonotic agent *Streptococcus suis*. Future Microbiology 7, 259–279. doi:10.2217/fmb.11.149
Fittipaldi, N., Takamatsu, D., Domínguez-Punaro, M. de la C., Lecours, M.-P., Montpetit, D., Osaki, M., Sekizaki, T., Gottschalk, M., 2010. Mutations in the gene encoding the ancillary pilin subunit of the *Streptococcus suis* srtF cluster result in pili

formed by the major subunit only. PLoS ONE 5, e8426. doi:10.1371/journal.pone.0008426
Fittipaldi, N., Xu, J., Lacouture, S., Tharavichitkul, P., Osaki, M., Sekizaki, T., Takamatsu, D., Gottschalk, M., 2011. Lineage and virulence of *Streptococcus suis* serotype 2 isolates from North America. Emerging Infect. Dis. 17, 2239–2244. doi:10.3201/eid1712.110609
Food and Drug Administration Center for Veterinary Medicine, 2003. Evaluating the safety of antimicrobial new animal drugs with regard to their microbiological effects on bacteria of human health concern. Guidance for Industry #152 1–36.
Foster, N., Staats, J.J., Chengappa, M.M., 1994. Isolation, characterization and protection studies in mice of a streptomycin-dependent mutant of *Streptococcus suis* type 1/2. Vet Res Commun 18, 155–163. doi:10.1007/BF01839263
Friedman, J., Hastie, T., Tibshirani, R., 2010. Regularization Paths for Generalized Linear Models via Coordinate Descent. Journal of Statistical Software 33, 1–22.
Frost, L.S., Leplae, R., Summers, A.O., Toussaint, A., 2005. Mobile genetic elements: the agents of open source evolution. Nat Rev Microbiol 3, 722–732. doi:10.1038/nrmicro1235
Galina, L., Collins, J.E., Pijoan, C., 1992. Porcine *Streptococcus suis* in Minnesota. Journal of Veterinary Diagnostic Investigation 4, 195–196. doi:10.1177/104063879200400216
Galina, L., Vecht, U., Wisselink, H.J., Pijoan, C., 1996. Prevalence of various phenotypes of *Streptococcus suis* isolated from swine in the U.S.A. based on the presence of muraminidase-released protein and extracellular factor. Can. J. Vet. Res. 60, 72–74.
Galperin, M.Y., Makarova, K.S., Wolf, Y.I., Koonin, E.V., 2015. Expanded microbial genome coverage and improved protein family annotation in the COG database. Nucleic Acids Res 43, D261-269. doi:10.1093/nar/gku1223
Gao, X.-Y., Zhi, X.-Y., Li, H.-W., Klenk, H.-P., Li, W.-J., 2014. Comparative genomics of the bacterial genus *Streptococcus* illuminates evolutionary implications of species groups. PLoS ONE 9, e101229. doi:10.1371/journal.pone.0101229
Geng, H., Zhu, L., Yuan, Y., Zhang, W., Li, W., Wang, J., Zheng, Y., Wei, K., Cao, W., Wang, H., Jiang, Y., 2008. Identification and characterization of novel immunogenic proteins of *Streptococcus suis* serotype 2. J. Proteome Res. 7, 4132–4142. doi:10.1021/pr800196v
Gérard Guédon, Virginie Libante, Charles Coluzzi, Sophie Payot, Nathalie Leblond-Bourget, 2017. The obscure world of integrative and mobilizable elements, highly widespread elements that pirate bacterial conjugative systems. Genes 8, 337. doi:10.3390/genes8110337
Gerber, J., Nau, R., 2010. Mechanisms of injury in bacterial meningitis. Curr. Opin. Neurol. 23, 312–318. doi:10.1097/WCO.0b013e32833950dd
Gobbetti, M., Calasso, M., 2014. STREPTOCOCCUS | Introduction, in: Encyclopedia of Food Microbiology. Elsevier, pp. 535–553. doi:10.1016/B978-0-12-384730-0.00324-4

Gottschalk, Marcelo, Higgins, R., Jacques, M., Beaudoin, M., Henrichsen, J., 1991. Isolation and characterization of *Streptococcus suis* capsular types 9-22. Journal of Veterinary Diagnostic Investigation 3, 60–65. doi:10.1177/104063879100300113

Gottschalk, M., Higgins, R., Jacques, M., Beaudoin, M., Henrichsen, J., 1991. Characterization of six new capsular types (23 through 28) of *Streptococcus suis*. J. Clin. Microbiol. 29, 2590–2594.

Gottschalk, M., Higgins, R., Jacques, M., Mittal, K.R., Henrichsen, J., 1989. Description of 14 new capsular types of *Streptococcus suis*. J. Clin. Microbiol. 27, 2633–2636.

Gottschalk, M., Higgins, R., Quessy, S., 1999. Dilemma of the virulence of *Streptococcus suis* strains. J. Clin. Microbiol. 37, 4202–4203.

Gottschalk, M., Lacouture, S., 2015. Canada: Distribution of *Streptococcus suis* (from 2012 to 2014) and Actinobacillus pleuropneumoniae (from 2011 to 2014) serotypes isolated from diseased pigs. Can. Vet. J. 56, 1093–1094.

Gottschalk, M., Lacouture, S., Bonifait, L., Roy, D., Fittipaldi, N., Grenier, D., 2013. Characterization of *Streptococcus suis* isolates recovered between 2008 and 2011 from diseased pigs in Québec, Canada. Vet. Microbiol. 162, 819–825. doi:10.1016/j.vetmic.2012.10.028

Gottschalk, M., Lebrun, A., Wisselink, H., Dubreuil, J.D., Smith, H., Vecht, U., 1998. Production of virulence-related proteins by Canadian strains of *Streptococcus suis* capsular type 2. Can. J. Vet. Res. 62, 75–79.

Gottschalk, M., Segura, M., 2019. Streptococcosis, in: Zimmerman, J.J., Karriker, L.A., Ramirez, A., Schwartz, K.J., Stevenson, G.W., Zhang, J. (Eds.), Diseases of Swine. Wiley, pp. 934–950. doi:10.1002/9781119350927.ch61

Gottschalk, M., Segura, M., 2000. The pathogenesis of the meningitis caused by *Streptococcus suis*: the unresolved questions. Vet. Microbiol. 76, 259–272.

Gottschalk, M., Segura, M., Xu, J., 2007. *Streptococcus suis* infections in humans: the Chinese experience and the situation in North America. Anim Health Res Rev 8, 29–45. doi:10.1017/S1466252307001247

Gottschalk, M., Xu, J., Lecours, M.-P., Grenier, D., Fittipaldi, N., Segura, M., 2010. *Streptococcus suis* infections in humans: what is the prognosis for Western countries? (Part II). Clinical Microbiology Newsletter 32, 97–102. doi:10.1016/j.clinmicnews.2010.06.001

Goyette-Desjardins, G., Auger, J.-P., Dolbec, D., Vinogradov, E., Okura, M., Takamatsu, D., Van Calsteren, M.-R., Gottschalk, M., Segura, M., 2020. Comparative study of immunogenic properties of purified capsular polysaccharides from *Streptococcus suis* serotypes 3, 7, 8, and 9: the serotype 3 polysaccharide induces an opsonizing IgG response. Infect Immun 88, e00377-20. doi:10.1128/IAI.00377-20

Goyette-Desjardins, G., Auger, J.-P., Xu, J., Segura, M., Gottschalk, M., 2014. *Streptococcus suis*, an important pig pathogen and emerging zoonotic agent-an update on the worldwide distribution based on serotyping and sequence typing. Emerg Microbes Infect 3, e45. doi:10.1038/emi.2014.45

Goyette-Desjardins, G., Calzas, C., Shiao, T.C., Neubauer, A., Kempker, J., Roy, R., Gottschalk, M., Segura, M., 2016. Protection against *Streptococcus suis* serotype 2 infection using a capsular polysaccharide glycoconjugate vaccine. Infect Immun 84, 2059–2075. doi:10.1128/IAI.00139-16

Goyette-Desjardins, G., Lacouture, S., Auger, J.-P., Roy, R., Gottschalk, M., Segura, M., 2019. Characterization and protective activity of monoclonal antibodies directed against *Streptococcus suis* serotype 2 capsular polysaccharide obtained using a glycoconjugate. Pathogens 8, 139. doi:10.3390/pathogens8030139

Guo, G., Du, D., Yu, Y., Zhang, Y., Qian, Y., Zhang, W., 2020. Pan-genome analysis of *Streptococcus suis* serotype 2 revealed genomic diversity among strains of different virulence. Transbound Emerg Dis 68, 637–647. doi:10.1111/tbed.13725

Gurevich, A., Saveliev, V., Vyahhi, N., Tesler, G., 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29, 1072–1075. doi:10.1093/bioinformatics/btt086

Gurung, M., Tamang, M.D., Moon, D.C., Kim, S.-R., Jeong, J.-H., Jang, G.-C., Jung, S.-C., Park, Y.-H., Lim, S.-K., 2015. Molecular basis of resistance to selected antimicrobial agents in the emerging zoonotic pathogen *Streptococcus suis*. J. Clin. Microbiol. 53, 2332–2336. doi:10.1128/JCM.00123-15

Hadjirin, N.F., Miller, E.L., Murray, G.G.R., Yen, P.L.K., Phuc, H.D., Wileman, T.M., Hernandez-Garcia, J., Williamson, S.M., Parkhill, J., Maskell, D.J., Zhou, R., Fittipaldi, N., Gottschalk, M., Tucker, A.W. (Dan), Hoa, N.T., Welch, J.J., Weinert, L.A., 2020. A comprehensive portrait of antimicrobial resistance in the zoonotic pathogen *Streptococcus suis* (preprint). Microbiology. doi:10.1101/2020.05.05.078493

Haesebrouck, F., Pasmans, F., Chiers, K., Maes, D., Ducatelle, R., Decostere, A., 2004. Efficacy of vaccines against bacterial diseases in swine: what can we expect? Vet. Microbiol. 100, 255–268. doi:10.1016/j.vetmic.2004.03.002

Hammerschmidt, S., Wolff, S., Hocke, A., Rosseau, S., Müller, E., Rohde, M., 2005. Illustration of pneumococcal polysaccharide capsule during adherence and invasion of epithelial cells. Infect Immun 73, 4653–4667. doi:10.1128/IAI.73.8.4653-4667.2005

Harel, J., Higgins, R., Gottschalk, M., Bigras-Poulin, M., 1994. Genomic relatedness among reference strains of different *Streptococcus suis* serotypes. Can J Vet Res 58, 259–262.

Hatrongjit, R., Fittipaldi, N., Gottschalk, M., Kerdsin, A., 2020. Tools for molecular epidemiology of *Streptococcus suis*. Pathogens 9. doi:10.3390/pathogens9020081

Hatrongjit, R., Kerdsin, A., Gottschalk, M., Takeuchi, D., Hamada, S., Oishi, K., Akeda, Y., 2015. First human case report of sepsis due to infection with *Streptococcus suis* serotype 31 in Thailand. BMC Infect. Dis. 15, 392. doi:10.1186/s12879-015-1136-0

Hayer, S.S., Rovira, A., Olsen, K., Johnson, T.J., Vannucci, F., Rendahl, A., Perez, A., Alvarez, J., 2020. Prevalence and time trend analysis of antimicrobial resistance in respiratory bacterial pathogens collected from diseased pigs in USA between

2006–2016. Research in Veterinary Science 128, 135–144. doi:10.1016/j.rvsc.2019.11.010

Hernandez-Garcia, J., Wang, J., Restif, O., Holmes, M.A., Mather, A.E., Weinert, L.A., Wileman, T.M., Thomson, J.R., Langford, P.R., Wren, B.W., Rycroft, A., Maskell, D.J., Tucker, A.W., 2017. Patterns of antimicrobial resistance in *Streptococcus suis* isolates from pigs with or without streptococcal disease in England between 2009 and 2014. Veterinary Microbiology 207, 117–124. doi:10.1016/j.vetmic.2017.06.002

Higgins, R., Gottschalk, M., 2001. Distribution of *Streptococcus suis* capsular types in 2000. Can. Vet. J. 42, 223.

Higgins, R., Gottschalk, M., Beaudoin, M., Rawluk, S.A., 1992. Distribution of *Streptococcus suis* capsular types in Quebec and western Canada. Can. Vet. J. 33, 27–30.

Higgins, R., Gottschalk, M., Boudreau, M., Lebrun, A., Henrichsen, J., 1995. Description of six new capsular types (29–34) of *Streptococcus suis*. Journal of Veterinary Diagnostic Investigation 7, 405–406. doi:10.1177/104063879500700322

Hill, J.E., Gottschalk, M., Brousseau, R., Harel, J., Hemmingsen, S.M., Goh, S.H., 2005. Biochemical analysis, cpn60 and 16S rDNA sequence data indicate that *Streptococcus suis* serotypes 32 and 34, isolated from pigs, are *Streptococcus orisratti*. Vet. Microbiol. 107, 63–69. doi:10.1016/j.vetmic.2005.01.003

Hoa, N.T., Chieu, T.T.B., Nga, T.T.T., Dung, N.V., Campbell, J., Anh, P.H., Huu Tho, H., Van Vinh Chau, N., Bryant, J.E., Hien, T.T., Farrar, J., Schultsz, C., 2011. Slaughterhouse pigs are a major reservoir of *Streptococcus suis* serotype 2 capable of causing human infection in Southern Vietnam. PLoS ONE 6, e17943. doi:10.1371/journal.pone.0017943

Holden, M.T.G., Hauser, H., Sanders, M., Ngo, T.H., Cherevach, I., Cronin, A., Goodhead, I., Mungall, K., Quail, M.A., Price, C., Rabbinowitsch, E., Sharp, S., Croucher, N.J., Chieu, T.B., Thi Hoang Mai, N., Diep, T.S., Chinh, N.T., Kehoe, M., Leigh, J.A., Ward, P.N., Dowson, C.G., Whatmore, A.M., Chanter, N., Iversen, P., Gottschalk, M., Slater, J.D., Smith, H.E., Spratt, B.G., Xu, J., Ye, C., Bentley, S., Barrell, B.G., Schultsz, C., Maskell, D.J., Parkhill, J., 2009. Rapid evolution of virulence and drug resistance in the emerging zoonotic pathogen *Streptococcus suis*. PLoS ONE 4, e6072. doi:10.1371/journal.pone.0006072

Holt, M.E., Enright, M.R., Alexander, T.J., 1990. Immunisation of pigs with killed cultures of *Streptococcus suis* type 2. Res Vet Sci 48, 23–27.

Holt, M.E., Enright, M.R., Alexander, T.J., 1988. Immunisation of pigs with live cultures of *Streptococcus suis* type 2. Res Vet Sci 45, 349–352.

Hopkins, D., Poljak, Z., Farzan, A., Friendship, R., 2019. Field studies evaluating the direct, indirect, total, and overall efficacy of *Streptococcus suis* autogenous vaccine in nursery pigs. Can. Vet. J. 60, 386–390.

Huang, J., Liu, X., Chen, H., Chen, L., Gao, X., Pan, Z., Wang, J., Lu, C., Yao, H., Wang, L., Wu, Z., 2019. Identification of six novel capsular polysaccharide loci (NCL) from *Streptococcus suis* multidrug resistant non-typeable strains and the

pathogenic characteristic of strains carrying new NCLs. Transbound Emerg Dis 66, 995–1003. doi:10.1111/tbed.13123

Huang, J., Ma, J., Shang, K., Hu, X., Liang, Y., Li, D., Wu, Z., Dai, L., Chen, L., Wang, L., 2016. Evolution and diversity of the antimicrobial resistance associated mobilome in *Streptococcus suis*: a probable mobile genetic elements reservoir for other streptococci. Front Cell Infect Microbiol 6, 118. doi:10.3389/fcimb.2016.00118

Hughes, J.M., Wilson, M.E., Wertheim, H.F.L., Nghia, H.D.T., Taylor, W., Schultsz, C., 2009. *Streptococcus suis:* an emerging human pathogen. Clinical Infectious Diseases 48, 617–625. doi:10.1086/596763

Huong, V.T.L., Ha, N., Huy, N.T., Horby, P., Nghia, H.D.T., Thiem, V.D., Zhu, X., Hoa, N.T., Hien, T.T., Zamora, J., Schultsz, C., Wertheim, H.F.L., Hirayama, K., 2014. Epidemiology, clinical manifestations, and outcomes of *Streptococcus suis* infection in humans. Emerging Infect. Dis. 20, 1105–1114. doi:10.3201/eid2007.131594

Inouye, M., Dashnow, H., Raven, L.-A., Schultz, M.B., Pope, B.J., Tomita, T., Zobel, J., Holt, K.E., 2014. SRST2: rapid genomic surveillance for public health and hospital microbiology labs. Genome Med 6, 90–105. doi:10.1186/s13073-014-0090-6

Ishida, S., Tien, L.H.T., Osawa, R., Tohya, M., Nomoto, R., Kawamura, Y., Takahashi, T., Kikuchi, N., Kikuchi, K., Sekizaki, T., 2014. Development of an appropriate PCR system for the reclassification of *Streptococcus suis*. Journal of Microbiological Methods 107, 66–70. doi:10.1016/j.mimet.2014.09.003

Jackson, R.W., Vinatzer, B., Arnold, D.L., Dorus, S., Murillo, J., 2011. The influence of the accessory genome on bacterial pathogen evolution. Mobile Genetic Elements 1, 55–65. doi:10.4161/mge.1.1.16432

Jacobs, A.A., Loeffen, P.L., van den Berg, A.J., Storm, P.K., 1994. Identification, purification, and characterization of a thiol-activated hemolysin (suilysin) of *Streptococcus suis*. Infect Immun 62, 1742–1748. doi:10.1128/iai.62.5.1742-1748.1994

Jacobs, A.A.C., van den Berg, A.J.G., Loeffen, P.L.W., 1996. Protection of experimentally infected pigs by suilysin, the thiol-activated haemolysin of *Streptococcus suis*. Veterinary Record 139, 225–228. doi:10.1136/vr.139.10.225

Jiang, X., Yang, Y., Zhou, J., Liu, H., Liao, X., Luo, J., Li, X., Fang, W., 2019. Peptidyl isomerase PrsA is surface-associated on *Streptococcus suis* and offers cross-protection against serotype 9 strain. FEMS Microbiology Letters 366, fnz002. doi:10.1093/femsle/fnz002

Jiang, X., Yang, Y., Zhou, J., Zhu, L., Gu, Y., Zhang, X., Li, X., Fang, W., 2016. Roles of the putative type IV-like secretion system key component VirD4 and PrsA in pathogenesis of *Streptococcus suis* type 2. Front Cell Infect Microbiol 6, 172. doi:10.3389/fcimb.2016.00172

Johansson, M.H.K., Bortolaia, V., Tansirichaiya, S., Aarestrup, F.M., Roberts, A.P., Petersen, T.N., 2021. Detection of mobile genetic elements associated with antibiotic resistance in *Salmonella enterica* using a newly developed web tool:

MobileElementFinder. J Antimicrob Chemother 76, 101–109. doi:10.1093/jac/dkaa390
Jolley, K.A., Bray, J.E., Maiden, M.C.J., 2018. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. Wellcome Open Research 3, 124. doi:10.12688/wellcomeopenres.14826.1
Jolley, K.A., Maiden, M.C., 2010. BIGSdb: Scalable analysis of bacterial genome variation at the population level. BMC Bioinformatics 11, 595–605. doi:10.1186/1471-2105-11-595
Kassambara, A., Mundt, F., 2017. factoextra: extract and visualize the results of multivariate data analyses.
Kebede, M., Chengappa, M.M., Stuart, J.G., 1990. Isolation and characterization of temperature-sensitive mutants of *Streptococcus suis*: Efficacy trial of the mutant vaccine in mice. Veterinary Microbiology 22, 249–257. doi:10.1016/0378-1135(90)90112-9
Kerdsin, A., Akeda, Y., Hatrongjit, R., Detchawna, U., Sekizaki, T., Hamada, S., Gottschalk, M., Oishi, K., 2014. *Streptococcus suis* serotyping by a new multiplex PCR. Journal of Medical Microbiology 63, 824–830. doi:10.1099/jmm.0.069757-0
Kerdsin, A., Akeda, Y., Takeuchi, D., Dejsirilert, S., Gottschalk, M., Oishi, K., 2018. Genotypic diversity of *Streptococcus suis* strains isolated from humans in Thailand. Eur. J. Clin. Microbiol. Infect. Dis. 37, 917–925. doi:10.1007/s10096-018-3208-8
Kerdsin, A., Hatrongjit, R., Gottschalk, M., Takeuchi, D., Hamada, S., Akeda, Y., Oishi, K., 2017. Emergence of *Streptococcus suis* serotype 9 infection in humans. Journal of Microbiology, Immunology and Infection 50, 545–546. doi:10.1016/j.jmii.2015.06.011
Kilian, M., Tettelin, H., 2019. Identification of virulence-associated properties by comparative genome analysis of *Streptococcus pneumoniae*, *S. pseudopneumoniae*, *S. mitis*, three *S. oralis* subspecies, and *S. infantis*. mBio 10, e01985-19. doi:10.1128/mBio.01985-19
Kim, K.S., 2008. Mechanisms of microbial traversal of the blood-brain barrier. Nat. Rev. Microbiol. 6, 625–634. doi:10.1038/nrmicro1952
King, S.J., Heath, P.J., Luque, I., Tarradas, C., Dowson, C.G., Whatmore, A.M., 2001. Distribution and genetic diversity of suilysin in *Streptococcus suis* isolated from different diseases of pigs and characterization of the genetic basis of suilysin absence. Infect. Immun. 69, 7572–7582. doi:10.1128/IAI.69.12.7572-7582.2001
King, S.J., Leigh, J.A., Heath, P.J., Luque, I., Tarradas, C., Dowson, C.G., Whatmore, A.M., 2002. Development of a multilocus sequence typing scheme for the pig pathogen *Streptococcus suis*: Identification of virulent clones and potential capsular serotype exchange. J. Clin. Microbiol. 40, 3671–3680.
Kong, D., Chen, Z., Wang, J., Lv, Q., Jiang, H., Zheng, Y., Xu, M., Zhou, X., Hao, H., Jiang, Y., 2017. Interaction of factor H-binding protein of *Streptococcus suis* with globotriaosylceramide promotes the development of meningitis. Virulence 8, 1290–1302. doi:10.1080/21505594.2017.1317426

Lachance, C., Gottschalk, M., Gerber, P.P., Lemire, P., Xu, J., Segura, M., 2013. Exacerbated type II interferon response drives hypervirulence and toxic shock by an emergent epidemic strain of *Streptococcus suis*. Infect Immun 81, 1928–1939. doi:10.1128/IAI.01317-12

Lakkitjaroen, N., Takamatsu, D., Okura, M., Sato, M., Osaki, M., Sekizaki, T., 2014. Capsule loss or death: The position of mutations among capsule genes sways the destiny of *Streptococcus suis*. FEMS Microbiol Lett 354, 46–54. doi:10.1111/1574-6968.12428

Lakkitjaroen, N., Takamatsu, D., Okura, M., Sato, M., Osaki, M., Sekizaki, T., 2011. Loss of capsule among *Streptococcus suis* isolates from porcine endocarditis and its biological significance. Journal of Medical Microbiology 60, 1669–1676. doi:10.1099/jmm.0.034686-0

Lalonde, M., Segura, M., Lacouture, S., Gottschalk, M., 2000. Interactions between *Streptococcus suis* serotype 2 and different epithelial cell lines. Microbiology 146, 1913–1921. doi:10.1099/00221287-146-8-1913

Lawrence, J.G., Retchless, A.C., 2009. The interplay of homologous recombination and horizontal gene transfer in bacterial speciation, in: Gogarten, M.B., Gogarten, J.P., Olendzenski, L.C. (Eds.), Horizontal Gene Transfer, Methods in Molecular Biology. Humana Press, Totowa, NJ, pp. 29–53. doi:10.1007/978-1-60327-853-9_3

Lê, S., Josse, J., Husson, F., 2008. FactoMineR: a package for multivariate analysis. Journal of Statistical Software 25, 1–18. doi:10.18637/jss.v025.i01

Lefébure, T., Stanhope, M.J., 2007. Evolution of the core and pan-genome of *Streptococcus*: positive selection, recombination, and genome composition. Genome Biol 8, R71. doi:10.1186/gb-2007-8-5-r71

Leuwerke, B., Wilson, K., Schulz, K., 2019. Control Strategies for *Streptococcus suis*. Webinar.

Li, D., Luo, R., Liu, C.-M., Leung, C.-M., Ting, H.-F., Sadakane, K., Yamashita, H., Lam, T.-W., 2016. MEGAHIT v1.0: a fast and scalable metagenome assembler driven by advanced methodologies and community practices. Methods 102, 3–11. doi:10.1016/j.ymeth.2016.02.020

Li, M., Wang, C., Feng, Y., Pan, X., Cheng, G., Wang, Jing, Ge, J., Zheng, F., Cao, M., Dong, Y., Liu, D., Wang, Jufang, Lin, Y., Du, H., Gao, G.F., Wang, X., Hu, F., Tang, J., 2008. SalK/SalR, a two-component signal transduction system, is essential for full virulence of highly invasive *Streptococcus suis* serotype 2. PLoS ONE 3, e2080. doi:10.1371/journal.pone.0002080

Li, Q., Fu, Y., Ma, C., He, Y., Yu, Y., Du, D., Yao, H., Lu, C., Zhang, W., 2017. The non-conserved region of MRP is involved in the virulence of *Streptococcus suis* serotype 2. Virulence 8, 1274–1289. doi:10.1080/21505594.2017.1313373

Li, W., Liu, L., Qiu, D., Chen, H., Zhou, R., 2010a. Identification of *Streptococcus suis* serotype 2 genes preferentially expressed in the natural host. International Journal of Medical Microbiology 300, 482–488. doi:10.1016/j.ijmm.2010.04.018

Li, W., Ye, C., Jing, H., Cui, Z., Bai, X., Jin, D., Zheng, H., Zhao, A., Xu, Y., Gottschalk, M., Xu, J., 2010b. *Streptococcus suis* outbreak investigation using multiple-locus

variable tandem repeat number analysis. Microbiology and Immunology 54, 380–388. doi:10.1111/j.1348-0421.2010.00228.x

Li, Y., Gottschalk, M., Esgleas, M., Lacouture, S., Dubreuil, J.D., Willson, P., Harel, J., 2007. Immunization with recombinant sao protein confers protection against *Streptococcus suis* infection. Clin Vaccine Immunol 14, 937–943. doi:10.1128/CVI.00046-07

Li, Y., Martinez, G., Gottschalk, M., Lacouture, S., Willson, P., Dubreuil, J.D., Jacques, M., Harel, J., 2006. Identification of a surface protein of *Streptococcus suis* and evaluation of its immunogenic and protective capacity in pigs. Infect Immun 74, 305–312. doi:10.1128/IAI.74.1.305-312.2006

Li, Z., Chang, P., Xu, J., Tan, C., Wang, X., Bei, W., Li, J., 2019. A *Streptococcus suis* live vaccine suppresses streptococcal toxic shock-like syndrome and provides sequence type-independent protection. The Journal of Infectious Diseases 219, 448–458. doi:10.1093/infdis/jiy512

Libante, V., Nombre, Y., Coluzzi, C., Staub, J., Guédon, G., Gottschalk, M., Teatero, S., Fittipaldi, N., Leblond-Bourget, N., Payot, S., 2019. Chromosomal conjugative and mobilizable elements in *Streptococcus suis*: major actors in the spreading of antimicrobial resistance and bacteriocin synthesis genes. Pathogens 9, 22. doi:10.3390/pathogens9010022

Liu, H., Lei, S., Jia, L., Xia, X., Sun, Y., Jiang, H., Zhu, R., Li, S., Qu, G., Gu, J., Sun, C., Feng, X., Han, W., Langford, P.R., Lei, L., 2021. *Streptococcus suis* serotype 2 enolase interaction with host brain microvascular endothelial cells and RPSA-induced apoptosis lead to loss of BBB integrity. Vet Res 52, 30. doi:10.1186/s13567-020-00887-6

Liu, L., Cheng, G., Wang, C., Pan, X., Cong, Y., Pan, Q., Wang, J., Zheng, F., Hu, F., Tang, J., 2009. Identification and experimental verification of protective antigens against *Streptococcus suis* serotype 2 based on genome sequence analysis. Curr Microbiol 58, 11–17. doi:10.1007/s00284-008-9258-x

Liu, M.Y., Romeo, T., 1997. The global regulator CsrA of *Escherichia coli* is a specific mRNA-binding protein. J. Bacteriol. 179, 4639–4642. doi:10.1128/jb.179.14.4639-4642.1997

Liu, Z., Zheng, H., Gottschalk, M., Bai, X., Lan, R., Ji, S., Liu, H., Xu, J., 2013. Development of multiplex PCR assays for the identification of the 33 serotypes of *Streptococcus suis*. PLoS ONE 8, e72070. doi:10.1371/journal.pone.0072070

Lowe, J., Gauger, P., Harmon, K., Zhang, J., Connor, J., Yeske, P., Loula, T., Levis, I., Dufresne, L., Main, R., 2014. Role of transportation in spread of porcine epidemic diarrhea virus infection, United States. Emerging Infect. Dis. 20, 872–874. doi:10.3201/eid2005.131628

Lugsomya, K., Yindee, J., Niyomtham, W., Tribuddharat, C., Tummaruk, P., Hampson, D.J., Prapasarakul, N., 2018. Antimicrobial resistance in commensal *Escherichia coli* isolated from pigs and pork derived from farms either routinely using or not using in-feed antimicrobials. Microb Drug Resist 24, 1054–1066. doi:10.1089/mdr.2018.0154

Lun, S., Perez-Casal, J., Connor, W., Willson, P.J., 2003. Role of suilysin in pathogenesis of *Streptococcus suis* capsular serotype 2. Microb Pathog 34, 27–37. doi:10.1016/s0882-4010(02)00192-4

Lun, Z.-R., Wang, Q.-P., Chen, X.-G., Li, A.-X., Zhu, X.-Q., 2007. *Streptococcus suis*: an emerging zoonotic pathogen. Lancet Infect Dis 7, 201–209. doi:10.1016/S1473-3099(07)70001-4

Luque, I., Blume, V., Borge, C., Vela, A.I., Perea, J.A., Márquez, J.M., Fernández-Garayzábal, J.F., Tarradas, C., 2010. Genetic analysis of *Streptococcus suis* isolates recovered from diseased and healthy carrier pigs at different stages of production on a pig farm. The Veterinary Journal 186, 396–398. doi:10.1016/j.tvjl.2009.09.005

Lütticken, R., Temme, N., Hahn, G., Bartelheimer, E.W., 1986. Meningitis caused by *Streptococcus suis*: case report and review of the literature. Infection 14, 181–185. doi:10.1007/BF01645260

Madsen, L.W., Svensmark, B., Elvestad, K., Aalbaek, B., Jensen, H.E., 2002. *Streptococcus suis* serotype 2 infection in pigs: new diagnostic and pathogenetic aspects. Journal of Comparative Pathology 126, 57–65. doi:10.1053/jcpa.2001.0522

Maneerat, K., Yongkiettrakul, S., Kramomtong, I., Tongtawe, P., Tapchaisri, P., Luangsuk, P., Chaicumpa, W., Gottschalk, M., Srimanote, P., 2013. Virulence genes and genetic diversity of *Streptococcus suis* serotype 2 isolates from Thailand. Transbound Emerg Dis 60 Suppl 2, 69–79. doi:10.1111/tbed.12157

Marini, E., Palmieri, C., Magi, G., Facinelli, B., 2015. Recombination between *Streptococcus suis* ICESsu32457 and *Streptococcus agalactiae* ICESa2603 yields a hybrid ICE transferable to *Streptococcus pyogenes*. Veterinary Microbiology 178, 99–104. doi:10.1016/j.vetmic.2015.04.013

Marois, C., Le Devendec, L., Gottschalk, M., Kobisch, M., 2007. Detection and molecular typing of *Streptococcus suis* in tonsils from live pigs in France. Can. J. Vet. Res. 71, 14–22.

Marois, C., Le Devendec, L., Gottschalk, M., Kobisch, M., 2006. Molecular characterization of *Streptococcus suis* strains by 16S-23S intergenic spacer polymerase chain reaction and restriction fragment length polymorphism analysis. Can J Vet Res 70, 94–104.

Martinez, G., Harel, J., Lacouture, S., Gottschalk, M., 2002. Genetic diversity of *Streptococcus suis* serotypes 2 and 1/2 isolates recovered from carrier pigs in closed herds. Can. J. Vet. Res. 66, 240–248.

Matajira, C.E.C., Moreno, L.Z., Poor, A.P., Gomes, V.T.M., Dalmutt, A.C., Parra, B.M., Oliveira, C.H. de, Barbosa, M.R.F., Sato, M.I.Z., Calderaro, F.F., Moreno, A.M., 2019. *Streptococcus suis* in Brazil: genotypic, virulence, and resistance profiling of strains isolated from pigs between 2001 and 2016. Pathogens 9, 31. doi:10.3390/pathogens9010031

McGuckin, M.A., Thornton, D.J., Whitsett, J.A., 2015. Mucins and Mucus, in: Mucosal Immunology. Elsevier, pp. 231–250. doi:10.1016/B978-0-12-415847-4.00014-8

Medini, D., Donati, C., Tettelin, H., Masignani, V., Rappuoli, R., 2005. The microbial pan-genome. Curr. Opin. Genet. Dev. 15, 589–594. doi:10.1016/j.gde.2005.09.006

Meng, F., Wu, N.-H., Seitz, M., Herrler, G., Valentin-Weigand, P., 2016. Efficient suilysin-mediated invasion and apoptosis in porcine respiratory epithelial cells after streptococcal infection under air-liquid interface conditions. Sci Rep 6, 26748. doi:10.1038/srep26748

Meng, P., Lu, C., Zhang, Q., Lin, J., Chen, F., 2017. Exploring the genomic diversity and cariogenic differences of *Streptococcus mutans* strains through pan-genome and comparative genome analysis. Curr. Microbiol. 74, 1200–1209. doi:10.1007/s00284-017-1305-z

Messier, S., Lacouture, S., Gottschalk, M., 2008. Distribution of *Streptococcus suis* capsular types from 2001 to 2007. Can. Vet. J. 49, 461–462.

Mielke, P.W., Berry, K.J., 2002. Categorical Independence Tests for Large Sparse *R* - Way Contingency Tables. Perceptual and Motor Skills 95, 606–610. doi:10.2466/pms.2002.95.2.606

Mogollon, J.D., Pijoan, C., Murtaugh, M.P., Collins, J.E., Cleary, P.P., 1991. Identification of epidemic strains of *Streptococcus suis* by genomic fingerprinting. J. Clin. Microbiol. 29, 782–787. doi:10.1128/JCM.29.4.782-787.1991

Morales, B., Ruiz, Á., Lacouture, S., Gottschalk, M., 2015. Clonal distribution of *Streptococcus suis* isolated from diseased pigs in the central region of Chile. Can. J. Vet. Res. 79, 343–346.

Mwaniki, C.G., Robertson, I.D., Trott, D.J., Atyeo, R.F., Lee, B.J., Hampson, D.J., 1994. Clonal analysis and virulence of Australian isolates of *Streptococcus suis* type 2. Epidemiol. Infect. 113, 321–334. doi:10.1017/s095026880005175x

Nelson, A.L., Roche, A.M., Gould, J.M., Chim, K., Ratner, A.J., Weiser, J.N., 2007. Capsule enhances pneumococcal colonization by limiting mucus-mediated clearance. Infect Immun 75, 83–90. doi:10.1128/IAI.01475-06

Nelson, M.I., Lemey, P., Tan, Y., Vincent, A., Lam, T.T.-Y., Detmer, S., Viboud, C., Suchard, M.A., Rambaut, A., Holmes, E.C., Gramer, M., 2011. Spatial dynamics of human-origin H1 influenza A virus in North American swine. PLoS Pathog. 7, e1002077. doi:10.1371/journal.ppat.1002077

Nicholson, T.L., Waack, U., Anderson, T.K., Bayles, D.O., Zaia, S.R., Goertz, I., Eppinger, M., Hau, S.J., Brockmeier, S.L., Shore, S.M., 2020. Comparative virulence and genomic analysis of *Streptococcus suis* isolates. Front Microbiol 11, 620843. doi:10.3389/fmicb.2020.620843

Nomoto, R., Maruyama, F., Ishida, S., Tohya, M., Sekizaki, T., Osawa, R., 2015. Reappraisal of the taxonomy of *Streptococcus suis* serotypes 20, 22 and 26: *Streptococcus parasuis* sp. nov. International Journal of Systematic and Evolutionary Microbiology 65, 438–443. doi:10.1099/ijs.0.067116-0

Novick, S., Shagan, M., Blau, K., Lifshitz, S., Givon-Lavi, N., Grossman, N., Bodner, L., Dagan, R., Nebenzahl, Y.M., 2017. Adhesion and invasion of *Streptococcus pneumoniae* to primary and secondary respiratory epithelial cells. Molecular Medicine Reports 15, 65–74. doi:10.3892/mmr.2016.5996

Obradovic, M.R., Segura, M., Segalés, J., Gottschalk, M., 2021. Review of the speculative role of co-infections in *Streptococcus suis*-associated diseases in pigs. Vet Res 52, 49. doi:10.1186/s13567-021-00918-w

O'Dea, M.A., Laird, T., Abraham, R., Jordan, D., Lugsomya, K., Fitt, L., Gottschalk, M., Truswell, A., Abraham, S., 2018. Examination of Australian *Streptococcus suis* isolates from clinically affected pigs in a global context and the genomic characterisation of ST1 as a predictor of virulence. Veterinary Microbiology 226, 31–40. doi:10.1016/j.vetmic.2018.10.010

Oh, S.-I., Jeon, A.B., Jung, B.-Y., Byun, J.-W., Gottschalk, M., Kim, A., Kim, J.W., Kim, H.-Y., 2017. Capsular serotypes, virulence-associated genes and antimicrobial susceptibility of *Streptococcus suis* isolates from pigs in Korea. The Journal of Veterinary Medical Science 79, 780–787. doi:10.1292/jvms.16-0514

Okura, M., Lachance, C., Osaki, M., Sekizaki, T., Maruyama, F., Nozawa, T., Nakagawa, I., Hamada, S., Rossignol, C., Gottschalk, M., Takamatsu, D., 2014. Development of a two-step multiplex PCR assay for typing of capsular polysaccharide synthesis gene clusters of *Streptococcus suis*. J. Clin. Microbiol. 52, 1714–1719. doi:10.1128/JCM.03411-13

Okwumabua, O., Staats, J., Chengappa, M.M., 1995. Detection of genomic heterogeneity in *Streptococcus suis* isolates by DNA restriction fragment length polymorphisms of rRNA genes (ribotyping). J Clin Microbiol 33, 968–972. doi:10.1128/jcm.33.4.968-972.1995

Oliveira, C., Faoro, H., Alves, L.R., Goldenberg, S., 2017. RNA-binding proteins and their role in the regulation of gene expression in *Trypanosoma cruzi* and *Saccharomyces cerevisiae*. Genet. Mol. Biol. 40, 22–30. doi:10.1590/1678-4685-GMB-2016-0258

Onishi, H., Sugawara, M., Okura, M., Osaki, M., Takamatsu, D., 2012. Prevalence of *Streptococcus Suis* genotypes in isolates from porcine endocarditis in East Japan. Journal of Veterinary Medical Science 74, 1681–1684. doi:10.1292/jvms.12-0301

Page, A.J., Cummins, C.A., Hunt, M., Wong, V.K., Reuter, S., Holden, M.T.G., Fookes, M., Falush, D., Keane, J.A., Parkhill, J., 2015. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics 31, 3691–3693. doi:10.1093/bioinformatics/btv421

Pallarés, F.J., Halbur, P.G., Schmitt, C.S., Roth, J.A., Opriessnig, T., Thomas, P.J., Kinyon, J.M., Murphy, D., Frank, D.E., Hoffman, L.J., 2003. Comparison of experimental models for *Streptococcus suis* infection of conventional pigs. Can. J. Vet. Res. 67, 225–228.

Pallarés, F.J., Schmitt, C.S., Roth, J.A., Evans, R.B., Kinyon, J.M., Halbur, P.G., 2004. Evaluation of a ceftiofur-washed whole cell *Streptococcus suis* bacterin in pigs. Can J Vet Res 68, 236–240.

Palmieri, C., Magi, G., Mingoia, M., Bagnarelli, P., Ripa, S., Varaldo, P.E., Facinelli, B., 2012. Characterization of a *Streptococcus suis* tet(O/W/32/O)-carrying element transferable to major streptococcal pathogens. Antimicrob Agents Chemother 56, 4697–4702. doi:10.1128/AAC.00629-12

Pan, Z., Ma, J., Dong, W., Song, W., Wang, K., Lu, C., Yao, H., 2015. Novel variant serotype of *Streptococcus suis* isolated from piglets with meningitis. Appl. Environ. Microbiol. 81, 976–985. doi:10.1128/AEM.02962-14

Partridge, S.R., Kwong, S.M., Firth, N., Jensen, S.O., 2018. Mobile genetic elements associated with antimicrobial resistance. Clin Microbiol Reviews 31, e00088-17. doi:10.1128/CMR.00088-17

Pasma, T., Davidson, S., Shaw, S.L., 2008. Idiopathic vesicular disease in swine in Manitoba. Can. Vet. J. 49, 84–85.

Peng, L., Lin, M., Huang, Z., Guo, S., Sun, H., Yang, X., 2020. Genetic analysis and pathogenicity of different sequence types of *Streptococcus suis* isolated from pigs in southern China. FEMS Microbiology Letters 367, fnaa049. doi:10.1093/femsle/fnaa049

Pérez-Sancho, M., Vela, A.I., García-Seco, T., González, S., Domínguez, L., Fernández-Garayzábal, J.F., 2017. Usefulness of MALDI-TOF MS as a diagnostic tool for the identification of *Streptococcus* species recovered from clinical specimens of pigs. PLoS ONE 12, e0170784. doi:10.1371/journal.pone.0170784

Petersen, L., Bollback, J.P., Dimmic, M., Hubisz, M., Nielsen, R., 2007. Genes under positive selection in *Escherichia coli*. Genome Research 17, 1336–1343. doi:10.1101/gr.6254707

Pian, Y., Li, X., Zheng, Y., Wu, X., Yuan, Y., Jiang, Y., 2016. Binding of human fibrinogen to MRP enhances *Streptococcus suis* survival in host blood in a αXβ2 integrin-dependent manner. Sci Rep 6, 26966. doi:10.1038/srep26966

Pig Health Today, 2019. Watch your Strep: keep your guard up for this evolving bacterium [WWW Document]. ThePigSite. URL https://www.thepigsite.com/news/2019/08/watch-your-strep-keep-your-guard-up-for-this-evolving-bacterium (accessed 12.1.20).

Prüfer, T.L., Rohde, J., Verspohl, J., Rohde, M., de Greeff, A., Willenborg, J., Valentin-Weigand, P., 2019. Molecular typing of *Streptococcus suis* strains isolated from diseased and healthy pigs between 1996-2016. PLoS ONE 14, e0210801. doi:10.1371/journal.pone.0210801

Qiu, X., Bai, X., Lan, R., Zheng, H., Xu, J., 2016. Novel capsular polysaccharide loci and new diagnostic tools for high-throughput capsular gene typing in *Streptococcus suis*. Appl Environ Microbiol 82, 7102–7112. doi:10.1128/AEM.02102-16

Quainoo, S., Coolen, J.P.M., van Hijum, S.A.F.T., Huynen, M.A., Melchers, W.J.G., van Schaik, W., Wertheim, H.F.L., 2017. Whole-genome sequencing of bacterial pathogens: the future of nosocomial outbreak analysis. Clin. Microbiol. Rev. 30, 1015–1063. doi:10.1128/CMR.00016-17

Quessy, S., Dubreuil, J.D., Caya, M., Létourneau, R., Higgins, R., 1994. Comparison of pig, rabbit and mouse IgG response to *Streptococcus suis* serotype 2 proteins and active immunization of mice against the infection. Can. J. Vet. Res. 58, 220–223.

R Core Team, 2017. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Rankin, D.J., Rocha, E.P.C., Brown, S.P., 2011. What traits are carried on mobile genetic elements, and why? Heredity (Edinb) 106, 1–10. doi:10.1038/hdy.2010.24

Rantsiou, K., Kathariou, S., Winkler, A., Skandamis, P., Saint-Cyr, M.J., Rouzeau-Szynalski, K., Amézquita, A., 2018. Next generation microbiological risk assessment: opportunities of whole genome sequencing (WGS) for foodborne pathogen surveillance, source tracking and risk assessment. Int. J. Food Microbiol. 287, 3–9. doi:10.1016/j.ijfoodmicro.2017.11.007

Rasko, D.A., Rosovitz, M.J., Myers, G.S.A., Mongodin, E.F., Fricke, W.F., Gajer, P., Crabtree, J., Sebaihia, M., Thomson, N.R., Chaudhuri, R., Henderson, I.R., Sperandio, V., Ravel, J., 2008. The pangenome structure of *Escherichia coli*: comparative genomic analysis of *E. coli* commensal and pathogenic isolates. J. Bacteriol. 190, 6881–6893. doi:10.1128/JB.00619-08

Reams, R.Y., Glickman, L.T., Harrington, D.D., Bowersock, T.L., Thacker, H.L., 1993. *Streptococcus suis* infection in swine: a retrospective study of 256 Cases. Part I. Epidemiologic factors and antibiotic susceptibility patterns. Journal of Veterinary Diagnostic Investigation 5, 363–367. doi:10.1177/104063879300500310

Reams, R.Y., Glickman, L.T., Harrington, D.D., Thacker, H.L., Bowersock, T.L., 1994. *Streptococcus suis* infection in swine: a retrospective study of 256 cases. Part II. Clinical signs, gross and microscopic lesions, and coexisting microorganisms. Journal of Veterinary Diagnostic Investigation 6, 326–334. doi:10.1177/104063879400600308

Reams, R.Y., Harrington, D.D., Glickman, L.T., Thacker, H.L., Bowersock, T.L., 1996. Multiple serotypes and strains of *Streptococcus suis* in naturally infected swine herds. J. Vet. Diagn. Invest. 8, 119–121. doi:10.1177/104063879600800121

Rehm, T., Baums, C.G., Strommenger, B., Beyerbach, M., Valentin-Weigand, P., Goethe, R., 2007. Amplified fragment length polymorphism of *Streptococcus suis* strains correlates with their profile of virulence-associated genes and clinical background. Journal of Medical Microbiology 56, 102–109. doi:10.1099/jmm.0.46616-0

Richards, V.P., Palmer, S.R., Pavinski Bitar, P.D., Qin, X., Weinstock, G.M., Highlander, S.K., Town, C.D., Burne, R.A., Stanhope, M.J., 2014. Phylogenomics and the dynamic genome evolution of the genus *Streptococcus*. Genome Biology and Evolution 6, 741–753. doi:10.1093/gbe/evu048

Rieckmann, K., Pendzialek, S.-M., Vahlenkamp, T., Baums, C.G., 2020. A critical review speculating on the protective efficacies of autogenous *Streptococcus suis* bacterins as used in Europe. Porcine Health Manag 6, 12. doi:10.1186/s40813-020-00150-6

Rieckmann, K., Seydel, A., Szewczyk, K., Klimke, K., Rungelrath, V., Baums, C.G., 2018. *Streptococcus suis cps7*: an emerging virulent sequence type (ST29) shows a distinct, IgM-determined pattern of bacterial survival in blood of piglets during the early adaptive immune response after weaning. Vet Res 49, 48. doi:10.1186/s13567-018-0544-8

Robertson, I.D., Blackmore, D.K., 1989. Occupational exposure to *Streptococcus suis* type 2. Epidemiol. Infect. 103, 157–164.

Rossi, C.C., Pereira, M.F., Giambiagi-deMarval, M., 2020. Underrated *Staphylococcus* species and their role in antimicrobial resistance spreading. Genet. Mol. Biol. 43, e20190065. doi:10.1590/1678-4685-gmb-2019-0065

Rouli, L., Merhej, V., Fournier, P.-E., Raoult, D., 2015. The bacterial pangenome as a new tool for analysing pathogenic bacteria. New Microbes New Infect 7, 72–85. doi:10.1016/j.nmni.2015.06.005

Roy, D., Grenier, D., Segura, M., Mathieu-Denoncourt, A., Gottschalk, M., 2016. Recruitment of Factor H to the *Streptococcus suis* cell surface is multifactorial. Pathogens 5, 47. doi:10.3390/pathogens5030047

Ruoff, K.L., Whiley, R.A., Beighton, D., 2003. *Streptococcus*, in: Murray, P.R., Baron, E.J., Jorgensen, J.H., Pfaller, M.A., Yolken, R.H. (Eds.), Manual of Clinical Microbiology. ASM Press, Washington, D.C., pp. 405–421.

Scherrer, S., Rosato, G., Spoerry Serrano, N., Stevens, M.J.A., Rademacher, F., Schrenzel, J., Gottschalk, M., Stephan, R., Peterhans, S., 2020. Population structure, genetic diversity and pathotypes of *Streptococcus suis* isolated during the last 13 years from diseased pigs in Switzerland. Vet Res 51, 85. doi:10.1186/s13567-020-00813-w

Schmid, S., O'Connor, M., Okwumabua, O., 2011. The pathogenicity island-like DNA segment associated with Chinese outbreak strain of *Streptococcus suis* serotype 2 is absent in the United States isolates. Int J Mol Epidemiol Genet 2, 56–60.

Schultsz, C., Jansen, E., Keijzers, W., Rothkamp, A., Duim, B., Wagenaar, J.A., van der Ende, A., 2012. Differences in the population structure of invasive *Streptococcus suis* strains isolated from pigs and from humans in the Netherlands. PLoS ONE 7, e33854. doi:10.1371/journal.pone.0033854

Schwerk, C., Papandreou, T., Schuhmann, D., Nickol, L., Borkowski, J., Steinmann, U., Quednau, N., Stump, C., Weiss, C., Berger, J., Wolburg, H., Claus, H., Vogel, U., Ishikawa, H., Tenenbaum, T., Schroten, H., 2012. Polar invasion and translocation of *Neisseria meningitidis* and *Streptococcus suis* in a novel human model of the blood-cerebrospinal fluid barrier. PLoS ONE 7, e30069. doi:10.1371/journal.pone.0030069

Seemann, T., 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30, 2068–2069. doi:10.1093/bioinformatics/btu153

Segers, R.P.A.M., Kenter, T., Haan, L.A.M., Jacobs, A.A.C., 1998. Characterisation of the gene encoding suilysin from *Streptococcus suis* and expression in field strains. FEMS Microbiology Letters 167, 255–261. doi:10.1111/j.1574-6968.1998.tb13236.x

Segura, M., 2015. *Streptococcus suis* vaccines: candidate antigens and progress. Expert Rev Vaccines 14, 1587–1608. doi:10.1586/14760584.2015.1101349

Segura, M., Aragon, V., Brockmeier, S.L., Gebhart, C., Greeff, A. de, Kerdsin, A., O'Dea, M.A., Okura, M., Saléry, M., Schultsz, C., Valentin-Weigand, P., Weinert, L.A., Wells, J.M., Gottschalk, M., 2020. Update on *Streptococcus suis* research and prevention in the era of antimicrobial restriction: 4th International Workshop on *S. suis*. Pathogens 9. doi:10.3390/pathogens9050374

Segura, M., Calzas, C., Grenier, D., Gottschalk, M., 2016. Initial steps of the pathogenesis of the infection caused by *Streptococcus suis*: fighting against nonspecific defenses. FEBS Lett 590, 3772–3799. doi:10.1002/1873-3468.12364

Segura, M., Fittipaldi, N., Calzas, C., Gottschalk, M., 2017. Critical *Streptococcus suis* virulence factors: are they all really critical? Trends in Microbiology 25, 585–599. doi:10.1016/j.tim.2017.02.005

Segura, M., Gottschalk, M., 2002. *Streptococcus suis* interactions with the murine macrophage cell line J774: adhesion and cytotoxicity. Infect Immun 70, 4312–4322. doi:10.1128/IAI.70.8.4312-4322.2002

Segura, M., Stankova, J., Gottschalk, M., 1999. Heat-killed *Streptococcus suis* capsular type 2 strains stimulate tumor necrosis factor alpha and interleukin-6 production by murine macrophages. Infect Immun 67, 4646–4654. doi:10.1128/IAI.67.9.4646-4654.1999

Seitz, M., Baums, C.G., Neis, C., Benga, L., Fulde, M., Rohde, M., Goethe, R., Valentin-Weigand, P., 2013. Subcytolytic effects of suilysin on interaction of *Streptococcus suis* with epithelial cells. Veterinary Microbiology 167, 584–591. doi:10.1016/j.vetmic.2013.09.010

Seitz, M., Beineke, A., Singpiel, A., Willenborg, J., Dutow, P., Goethe, R., Valentin-Weigand, P., Klos, A., Baums, C.G., 2014. Role of capsule and suilysin in mucosal infection of complement-deficient mice with *Streptococcus suis*. Infect Immun 82, 2460–2471. doi:10.1128/IAI.00080-14

Seitz, M., Valentin-Weigand, P., Willenborg, J., 2016. Use of antibiotics and antimicrobial resistance in veterinary medicine as exemplified by the swine pathogen *Streptococcus suis*. Curr. Top. Microbiol. Immunol. 398, 103–121. doi:10.1007/82_2016_506

Shelyakin, P.V., Bochkareva, O.O., Karan, A.A., Gelfand, M.S., 2019. Micro-evolution of three *Streptococcus* species: selection, antigenic variation, and horizontal gene inflow. BMC Evol Biol 19, 83. doi:10.1186/s12862-019-1403-6

Silva, L.M.G., Baums, C.G., Rehm, T., Wisselink, H.J., Goethe, R., Valentin-Weigand, P., 2006. Virulence-associated gene profiling of *Streptococcus suis* isolates by PCR. Vet. Microbiol. 115, 117–127. doi:10.1016/j.vetmic.2005.12.013

Smith, H.E., van Bruijnsvoort, L., Buijs, H., Wisselink, H.J., Smits, M.A., 1999. Rapid PCR test for *Streptococcus suis* serotype 7. FEMS Microbiol. Lett. 178, 265–270. doi:10.1111/j.1574-6968.1999.tb08686.x

Smith, H.E., Vecht, U., Wisselink, H.J., Stockhofe-Zurwieden, N., Biermann, Y., Smits, M.A., 1996. Mutants of *Streptococcus suis* types 1 and 2 impaired in expression of muramidase-released protein and extracellular protein induce disease in newborn germfree pigs. Infect. Immun. 64, 4409–4412.

Smith, M.R., 2017. Ternary: An R Package for Creating Ternary Plots. Zenodo.

Sourial, N., Wolfson, C., Zhu, B., Quail, J., Fletcher, J., Karunananthan, S., Bandeen-Roche, K., Béland, F., Bergman, H., 2010. Correspondence analysis is a useful tool to uncover the relationships among categorical variables. J Clin Epidemiol 63, 638–646. doi:10.1016/j.jclinepi.2009.08.008

Souvorov, A., Agarwala, R., Lipman, D.J., 2018. SKESA: strategic k-mer extension for scrupulous assemblies. Genome Biol 19, 153. doi:10.1186/s13059-018-1540-z

Staats, J.J., Feder, I., Okwumabua, O., Chengappa, M.M., 1997. *Streptococcus suis*: past and present. Vet. Res. Commun. 21, 381–407.

Stackebrandt, E., Goebel, B.M., 1994. Taxonomic note: a place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. International Journal of Systematic and Evolutionary Microbiology 44, 846–849. doi:10.1099/00207713-44-4-846

Su, Y., Yao, W., Perez-Gutierrez, O.N., Smidt, H., Zhu, W.-Y., 2008. Changes in abundance of *Lactobacillus spp.* and *Streptococcus suis* in the stomach, jejunum and ileum of piglets after weaning. FEMS Microbiology Ecology 66, 546–555. doi:10.1111/j.1574-6941.2008.00529.x

Sun, Y., Li, N., Zhang, J., Liu, H., Liu, J., Xia, X., Sun, C., Feng, X., Gu, J., Du, C., Han, W., Lei, L., 2016. Enolase of *Streptococcus suis* serotype 2 enhances blood–brain barrier permeability by inducing IL-8 release. Inflammation 39, 718–726. doi:10.1007/s10753-015-0298-7

Swildens, B., Stockhofe-Zurwieden, N., van der Meulen, J., Wisselink, H.J., Nielen, M., Niewold, T.A., 2004. Intestinal translocation of *Streptococcus suis* type 2 EF+ in pigs. Veterinary Microbiology 103, 29–33. doi:10.1016/j.vetmic.2004.06.010

Takamatsu, D., Nishino, H., Ishiji, T., Ishii, J., Osaki, M., Fittipaldi, N., Gottschalk, M., Tharavichitkul, P., Takai, S., Sekizaki, T., 2009. Genetic organization and preferential distribution of putative pilus gene clusters in *Streptococcus suis*. Vet. Microbiol. 138, 132–139. doi:10.1016/j.vetmic.2009.02.013

Takamatsu, D., Osaki, M., Tharavichitkul, P., Takai, S., Sekizaki, T., 2008a. Allelic variation and prevalence of serum opacity factor among the *Streptococcus suis* population. Journal of Medical Microbiology 57, 488–494. doi:10.1099/jmm.0.47755-0

Takamatsu, D., Wongsawan, K., Osaki, M., Nishino, H., Ishiji, T., Tharavichitkul, P., Khantawa, B., Fongcom, A., Takai, S., Sekizaki, T., 2008b. *Streptococcus suis* in Humans, Thailand. Emerg. Infect. Dis. 14, 181–183. doi:10.3201/eid1401.070568

Tan, M. -F., Tan, J., Zeng, Y. -B., Li, H. -Q., Yang, Q., Zhou, R., 2021. Antimicrobial resistance phenotypes and genotypes of *Streptococcus suis* isolated from clinically healthy pigs from 2017 to 2019 in Jiangxi Province, China. J. Appl. Microbiol. 130, 797–806. doi:10.1111/jam.14831

Tan, S., Dvorak, C.M.T., Estrada, A.A., Gebhart, C., Marthaler, D.G., Murtaugh, M.P., 2020. MinION sequencing of *Streptococcus suis* allows for functional characterization of bacteria by multilocus sequence typing and antimicrobial resistance profiling. J Microbiol Methods 169, 105817. doi:10.1016/j.mimet.2019.105817

Tang, J., Wang, C., Feng, Y., Yang, W., Song, H., Chen, Zhihai, Yu, H., Pan, X., Zhou, X., Wang, Huaru, Wu, B., Wang, Haili, Zhao, H., Lin, Y., Yue, J., Wu, Z., He, X., Gao, F., Khan, A.H., Wang, J., Zhao, G.-P., Wang, Y., Wang, X., Chen, Zhu, Gao, G.F., 2006. Streptococcal toxic shock syndrome caused by *Streptococcus suis* serotype 2. PLoS Med 3, e151. doi:10.1371/journal.pmed.0030151

Tenenbaum, T., Adam, R., Eggelnpöhlera, I., Matalon, D., Seibt, A., Novotny, G.E.K., Galla, H.-J., Schroten, H., 2005. Strain-dependent disruption of blood–cerebrospinal fluid barrier by *Streptoccocus suis* in vitro. FEMS Immunology & Medical Microbiology 44, 25–34. doi:10.1016/j.femsim.2004.12.006

Tenenbaum, T., Asmat, T.M., Seitz, M., Schroten, H., Schwerk, C., 2016. Biological activities of suilysin: role in *Streptococcus suis* pathogenesis. Future Microbiol 11, 941–954. doi:10.2217/fmb-2016-0028

Tenenbaum, T., Essmann, F., Adam, R., Seibt, A., Jänicke, R.U., Novotny, G.E.K., Galla, H.-J., Schroten, H., 2006. Cell death, caspase activation, and HMGB1 release of porcine choroid plexus epithelial cells during *Streptococcus suis* infection in vitro. Brain Research 1100, 1–12. doi:10.1016/j.brainres.2006.05.041

Tenenbaum, T., Papandreou, T., Gellrich, D., Friedrichs, U., Seibt, A., Adam, R., Wewer, C., Galla, H.-J., Schwerk, C., Schroten, H., 2009. Polar bacterial invasion and translocation of *Streptococcus suis* across the blood-cerebrospinal fluid barrier *in vitro*. Cellular Microbiology 11, 323–336. doi:10.1111/j.1462-5822.2008.01255.x

Thongkamkoon, P., Kiatyingangsulee, T., Gottschalk, M., 2017. Serotypes of *Streptococcus suis* isolated from healthy pigs in Phayao Province, Thailand. BMC Res Notes 10, 53. doi:10.1186/s13104-016-2354-2

Tibshirani, R., 1996. Regression Shrinkage and Selection via the Lasso. Journal of the Royal Statistical Society. Series B (Methodological) 58, 267–288.

Tien, L.H.T., Nishibori, T., Nishitani, Y., Nomoto, R., Osawa, R., 2013. Reappraisal of the taxonomy of *Streptococcus suis* serotypes 20, 22, 26, and 33 based on DNA–DNA homology and sodA and recN phylogenies. Vet. Microbiol. 162, 842–849. doi:10.1016/j.vetmic.2012.11.001

Tohya, M., Arai, S., Tomida, J., Watanabe, T., Kawamura, Y., Katsumi, M., Ushimizu, M., Ishida-Kuroki, K., Yoshizumi, M., Uzawa, Y., Iguchi, S., Yoshida, A., Kikuchi, K., Sekizaki, T., 2017. Defining the taxonomic status of *Streptococcus suis* serotype 33: the proposal for *Streptococcus ruminantium* sp. nov. Int. J. Syst. Evol. Microbiol. 67, 3660–3665. doi:10.1099/ijsem.0.002204

Torremorell, M., Calsamiglia, M., Pijoan, C., 1998. Colonization of suckling pigs by *Streptococcus suis* with particular reference to pathogenic serotype 2 strains. Can J Vet Res 62, 21–26.

Torremorell, M., Pijoan, C., 1998. Prolonged persistence of an epidemic *Streptococcus suis* strain in a closed pig population. Vet. Rec. 143, 394–395. doi:10.1136/vr.143.14.394

Tram, G., Jennings, M.P., Blackall, P.J., Atack, J.M., 2021. *Streptococcus suis* pathogenesis—A diverse array of virulence factors for a zoonotic lifestyle, in: Advances in Microbial Physiology. Elsevier, pp. 217–257. doi:10.1016/bs.ampbs.2020.12.002

Uelze, L., Grützke, J., Borowiak, M., Hammerl, J.A., Juraschek, K., Deneke, C., Tausch, S.H., Malorny, B., 2020. Typing methods based on whole genome sequencing data. One Health Outlook 2, 3. doi:10.1186/s42522-020-0010-1

Van Assche, E., Van Puyvelde, S., Vanderleyden, J., Steenackers, H.P., 2015. RNA-binding proteins involved in post-transcriptional regulation in bacteria. Front Microbiol 6, 141. doi:10.3389/fmicb.2015.00141

van Belkum, A., Tassios, P.T., Dijkshoorn, L., Haeggman, S., Cookson, B., Fry, N.K., Fussing, V., Green, J., Feil, E., Gerner-Smidt, P., Brisse, S., Struelens, M., European Society of Clinical Microbiology and Infectious Diseases (ESCMID) Study Group on Epidemiological Markers (ESGEM), 2007. Guidelines for the validation and application of typing methods for use in bacterial epidemiology. Clin. Microbiol. Infect. 13 Suppl 3, 1–46. doi:10.1111/j.1469-0691.2007.01786.x

Van Goethem, N., Descamps, T., Devleesschauwer, B., Roosens, N.H.C., Boon, N.A.M., Van Oyen, H., Robert, A., 2019. Status and potential of bacterial genomics for public health practice: a scoping review. Implement Sci 14, 79. doi:10.1186/s13012-019-0930-2

van Vliet, A.H.M., 2017. Use of pan-genome analysis for the identification of lineage-specific genes of *Helicobacter pylori*. FEMS Microbiology Letters 364, fnw296. doi:10.1093/femsle/fnw296

Vanier, G., Segura, M., Friedl, P., Lacouture, S., Gottschalk, M., 2004. Invasion of porcine brain microvascular endothelial cells by *Streptococcus suis* serotype 2. Infection and Immunity 72, 1441–1449. doi:10.1128/IAI.72.3.1441-1449.2004

Vanier, G., Segura, M., Lecours, M.-P., Grenier, D., Gottschalk, M., 2009. Porcine brain microvascular endothelial cell-derived interleukin-8 is first induced and then degraded by *Streptococcus suis*. Microbial Pathogenesis 46, 135–143. doi:10.1016/j.micpath.2008.11.004

Varela, N.P., Gadbois, P., Thibault, C., Gottschalk, M., Dick, P., Wilson, J., 2013. Antimicrobial resistance and prudent drug use for *Streptococcus suis*. Anim Health Res Rev 14, 68–77. doi:10.1017/S1466252313000029

Vecht, U., Wisselink, H.J., van Dijk, J.E., Smith, H.E., 1992. Virulence of *Streptococcus suis* type 2 strains in newborn germfree pigs depends on phenotype. Infect. Immun. 60, 550–556. doi:10.1128/iai.60.2.550-556.1992

Vela, A.I., Goyache, J., Tarradas, C., Luque, I., Mateos, A., Moreno, M.A., Borge, C., Perea, J.A., Dominguez, L., Fernandez-Garayzabal, J.F., 2003. Analysis of genetic diversity of *Streptococcus suis* clinical isolates from pigs in Spain by pulsed-field gel electrophoresis. Journal of Clinical Microbiology 41, 2498–2502. doi:10.1128/JCM.41.6.2498-2502.2003

von Wintersdorff, C.J.H., Penders, J., van Niekerk, J.M., Mills, N.D., Majumder, S., van Alphen, L.B., Savelkoul, P.H.M., Wolffs, P.F.G., 2016. Dissemination of antimicrobial resistance in microbial ecosystems through horizontal gene transfer. Front. Microbiol. 7, 173. doi:10.3389/fmicb.2016.00173

Vötsch, D., Willenborg, M., Weldearegay, Y.B., Valentin-Weigand, P., 2018. *Streptococcus suis* – the "two faces" of a pathobiont in the porcine respiratory tract. Front. Microbiol. 9, 480. doi:10.3389/fmicb.2018.00480

Vrancianu, C.O., Popa, L.I., Bleotu, C., Chifiriuc, M.C., 2020. Targeting plasmids to limit acquisition and transmission of antimicrobial resistance. Front Microbiol 11, 761. doi:10.3389/fmicb.2020.00761

Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q., Wortman, J., Young, S.K., Earl, A.M., 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS ONE 9, e112963. doi:10.1371/journal.pone.0112963

Wang, J., Kong, D., Zhang, S., Jiang, H., Zheng, Y., Zang, Y., Hao, H., Jiang, Y., 2015. Interaction of fibrinogen and muramidase-released protein promotes the development of *Streptococcus suis* meningitis. Front Microbiol 6, 1001. doi:10.3389/fmicb.2015.01001

Wang, K., Lu, C., 2007. Adhesion activity of glyceraldehyde-3-phosphate dehydrogenase in a Chinese *Streptococcus suis* type 2 strain. Berl Munch Tierarztl Wochenschr 120, 207–209.

Wang, Y., Li, D., Song, L., Liu, Y., He, T., Liu, H., Wu, C., Schwarz, S., Shen, J., 2013. First report of the multiresistance gene cfr in *Streptococcus suis*. Antimicrob Agents Chemother 57, 4061–4063. doi:10.1128/AAC.00713-13

Ward, P.N., Holden, M.T., Leigh, J.A., Lennard, N., Bignell, A., Barron, A., Clark, L., Quail, M.A., Woodward, J., Barrell, B.G., Egan, S.A., Field, T.R., Maskell, D., Kehoe, M., Dowson, C.G., Chanter, N., Whatmore, A.M., Bentley, S.D., Parkhill, J., 2009. Evidence for niche adaptation in the genome of the bovine pathogen *Streptococcus uberis*. BMC Genomics 10, 54. doi:10.1186/1471-2164-10-54

Warneboldt, F., Sander, S., Beineke, A., Valentin-Weigand, P., Kamphues, J., Baums, C., 2016. Clearance of *Streptococcus suis* in stomach contents of differently fed growing pigs. Pathogens 5, 56. doi:10.3390/pathogens5030056

Warnes, G.R., Bolker, B., Bonebakker, L., Gentleman, R., Liaw, W.H.A., Lumley, T., Maechler, M., Magnusson, A., Moeller, S., Schwartz, M., Venables, B., 2016. gplots: various R programming tools for plotting data.

Wei, Z., Li, R., Zhang, A., He, H., Hua, Y., Xia, J., Cai, X., Chen, H., Jin, M., 2009. Characterization of *Streptococcus suis* isolates from the diseased pigs in China between 2003 and 2007. Vet. Microbiol. 137, 196–201. doi:10.1016/j.vetmic.2008.12.015

Weinert, L.A., Chaudhuri, R.R., Wang, J., Peters, S.E., Corander, J., Jombart, T., Baig, A., Howell, K.J., Vehkala, M., Välimäki, N., Harris, D., Chieu, T.T.B., Van Vinh Chau, N., Campbell, J., Schultsz, C., Parkhill, J., Bentley, S.D., Langford, P.R., Rycroft, A.N., Wren, B.W., Farrar, J., Baker, S., Hoa, N.T., Holden, M.T.G., Tucker, A.W., Maskell, D.J., 2015. Genomic signatures of human and animal disease in the zoonotic pathogen *Streptococcus suis*. Nature Communications 6, 6740–6748. doi:10.1038/ncomms7740

Werinder, A., Aspán, A., Backhans, A., Sjölund, M., Guss, B., Jacobson, M., 2020. *Streptococcus suis* in Swedish grower pigs: occurrence, serotypes, and antimicrobial susceptibility. Acta Vet Scand 62, 36. doi:10.1186/s13028-020-00533-3

Westermann, A.J., Venturini, E., Sellin, M.E., Förstner, K.U., Hardt, W.-D., Vogel, J., 2019. The major RNA-binding protein ProQ impacts virulence gene expression in *Salmonella enterica* serovar Typhimurium. mBio 10, e02504-18. doi:10.1128/mBio.02504-18

Wichgers Schreur, P.J., van Weeghel, C., Rebel, J.M.J., Smits, M.A., van Putten, J.P.M., Smith, H.E., 2012. Lysozyme resistance in *Streptococcus suis* is highly variable and multifactorial. PLoS ONE 7, e36281. doi:10.1371/journal.pone.0036281

Wickham, H., 2017. tidyverse: easily install and load the "tidyverse."

Wickham, H., 2009. ggplot2: elegant graphics for data analysis. Springer-Verlag New York.

Wileman, T.M., Weinert, L.A., Howell, K.J., Wang, J., Peters, S.E., Williamson, S.M., Wells, J.M., Langford, P.R., Rycroft, A.N., Wren, B.W., Maskell, D.J., Tucker, A.W., 2019. Pathotyping the zoonotic pathogen *Streptococcus suis*: novel genetic markers to differentiate invasive disease-associated isolates from non-disease-associated isolates from England and Wales. J. Clin. Microbiol. 57, e01712-18. doi:10.1128/JCM.01712-18

Willemse, N., Howell, K.J., Weinert, L.A., Heuvelink, A., Pannekoek, Y., Wagenaar, J.A., Smith, H.E., van der Ende, A., Schultsz, C., 2016. An emerging zoonotic clone in the Netherlands provides clues to virulence and zoonotic potential of *Streptococcus suis*. Sci Rep 6, 28984. doi:10.1038/srep28984

Willemse, N., van der Ark, K.C.H., Stockhofe-Zurwieden, N., Smith, H., Picavet, D.I., van Solt-Smits, C., Wisselink, H.J., Schultsz, C., de Greeff, A., 2019. Clonal expansion of a virulent *Streptococcus suis* serotype 9 lineage distinguishable from carriage subpopulations. Sci Rep 9, 15429. doi:10.1038/s41598-019-51576-0

Williams, A.E., Blakemore, W.F., 1990. Pathogenesis of meningitis caused by *Streptococcus suis* type 2. Journal of Infectious Diseases 162, 474–481. doi:10.1093/infdis/162.2.474

Wisselink, H.J., Smith, H.E., Stockhofe-Zurwieden, N., Peperkamp, K., Vecht, U., 2000. Distribution of capsular types and production of muramidase-released protein (MRP) and extracellular factor (EF) of *Streptococcus suis* strains isolated from diseased pigs in seven European countries. Vet. Microbiol. 74, 237–248.

Wisselink, H.J., Stockhofe-Zurwieden, N., Hilgers, L.A.T., Smith, H.E., 2002. Assessment of protective efficacy of live and killed vaccines based on a non-encapsulated mutant of *Streptococcus suis* serotype 2. Veterinary Microbiology 84, 155–168. doi:10.1016/S0378-1135(01)00452-7

Wisselink, H.J., Vecht, U., Smith, H.E., Stockhofe-Zurwieden, N., 2001. Protection of pigs against challenge with virulent *Streptococcus suis* serotype 2 strains by a muramidase-released protein and extracellular factor vaccine. Veterinary Record 148, 473–477. doi:10.1136/vr.148.15.473

Wongsawan, K., Gottschalk, M., Tharavichitkul, P., 2015. Serotype- and virulence-associated gene profile of *Streptococcus suis* isolates from pig carcasses in Chiang Mai Province, Northern Thailand. J. Vet. Med. Sci. 77, 233–236. doi:10.1292/jvms.14-0380

Yang, Q.E., Walsh, T.R., 2017. Toxin-antitoxin systems and their role in disseminating and maintaining antimicrobial resistance. FEMS Microbiol Rev 41, 343–353. doi:10.1093/femsre/fux006

Yang, S., Rothman, R.E., 2004. PCR-based diagnostics for infectious diseases: uses, limitations, and future applications in acute-care settings. The Lancet Infectious Diseases 4, 337–348. doi:10.1016/S1473-3099(04)01044-8

Ye, C., Bai, X., Zhang, J., Jing, H., Zheng, H., Du, H., Cui, Z., Zhang, S., Jin, D., Xu, Y., Xiong, Y., Zhao, A., Luo, X., Sun, Q., Gottschalk, M., Xu, J., 2008. Spread of *Streptococcus suis* sequence type 7, China. Emerg. Infect. Dis. 14, 787–791. doi:10.3201/eid1405.070437

Ye, C., Zheng, H., Zhang, J., Jing, H., Wang, L., Xiong, Y., Wang, W., Zhou, Z., Sun, Q., Luo, X., Du, H., Gottschalk, M., Xu, J., 2009. Clinical, experimental, and genomic differences between intermediately pathogenic, highly pathogenic, and epidemic *Streptococcus suis*. J Infect Dis. 199, 97–107. doi:10.1086/594370

Yongkiettrakul, S., Maneerat, K., Arechanajan, B., Malila, Y., Srimanote, P., Gottschalk, M., Visessanguan, W., 2019. Antimicrobial susceptibility of *Streptococcus suis* isolated from diseased pigs, asymptomatic pigs, and human patients in Thailand. BMC Vet. Res. 15, 5. doi:10.1186/s12917-018-1732-5

Yoon, S.H., Park, Y.-K., Kim, J.F., 2015. PAIDB v2.0: exploration and analysis of pathogenicity and resistance islands. Nucleic Acids Res 43, D624-630. doi:10.1093/nar/gku985

Zhang, A., Chen, B., Mu, X., Li, R., Zheng, P., Zhao, Y., Chen, H., Jin, M., 2009. Identification and characterization of a novel protective antigen, Enolase of *Streptococcus suis* serotype 2. Vaccine 27, 1348–1353. doi:10.1016/j.vaccine.2008.12.047

Zhang, A., Yang, M., Hu, P., Wu, J., Chen, B., Hua, Y., Yu, J., Chen, H., Xiao, J., Jin, M., 2011. Comparative genomic analysis of *Streptococcus suis* reveals significant genomic diversity among different serotypes. BMC Genomics 12, 523–533. doi:10.1186/1471-2164-12-523

Zhang, B., Ku, X., Yu, X., Sun, Q., Wu, H., Chen, F., Zhang, X., Guo, L., Tang, X., He, Q., 2019. Prevalence and antimicrobial susceptibilities of bacterial pathogens in Chinese pig farms from 2013 to 2017. Sci Rep 9, 9908. doi:10.1038/s41598-019-45482-8

Zhang, C., Zhang, P., Wang, Y., Fu, L., Liu, L., Xu, D., Hou, Y., Li, Y., Fu, M., Wang, X., Wang, S., Ding, S., Shen, Z., 2020. Capsular serotypes, antimicrobial susceptibility, and the presence of transferable oxazolidinone resistance genes in *Streptococcus suis* isolated from healthy pigs in China. Veterinary Microbiology 247, 108750. doi:10.1016/j.vetmic.2020.108750

Zhang, W., Lu, C.P., 2007. Immunoproteomics of extracellular proteins of Chinese virulent strains of *Streptococcus suis* type 2. Proteomics 7, 4468–4476. doi:10.1002/pmic.200700294

Zhang, Y.-M., Shao, Z.-Q., Wang, J., Wang, L., Li, X., Wang, C., Tang, J., Pan, X., 2014. Prevalent distribution and conservation of *Streptococcus suis* Lmb protein and its protective capacity against the Chinese highly virulent strain infection. Microbiological Research 169, 395–401. doi:10.1016/j.micres.2013.09.007

Zheng, H., Du, P., Qiu, X., Kerdsin, A., Roy, D., Bai, X., Xu, J., Vela, A.I., Gottschalk, M., 2018. Genomic comparisons of *Streptococcus suis* serotype 9 strains

recovered from diseased pigs in Spain and Canada. Vet Res 49, 1. doi:10.1186/s13567-017-0498-2

Zheng, H., Ji, S., Lan, R., Liu, Z., Bai, X., Zhang, W., Gottschalk, M., Xu, J., 2014. Population analysis of *Streptococcus suis* isolates from slaughtered swine by use of minimum core genome sequence typing. Journal of Clinical Microbiology 52, 3568–3572. doi:10.1128/JCM.00536-14

Zheng, H., Ji, S., Liu, Z., Lan, R., Huang, Y., Bai, X., Gottschalk, M., Xu, J., 2015. Eight novel capsular polysaccharide synthesis gene loci identified in nontypeable *Streptococcus suis* isolates. Appl Environ Microbiol 81, 4111–4119. doi:10.1128/AEM.00315-15

Zheng, Han, Lan, R., Zheng, X., Cui, Z., Liu, Z., Bai, X., Ji, S., Gottschalk, M., Xu, J., 2014. Comparative genomic hybridization identifies virulence differences in *Streptococcus suis*. PLoS ONE 9, e87866. doi:10.1371/journal.pone.0087866

Zheng, H., Qiu, X., Roy, D., Segura, M., Du, P., Xu, J., Gottschalk, M., 2017. Genotyping and investigating capsular polysaccharide synthesis gene loci of non-serotypeable *Streptococcus suis* isolated from diseased pigs in Canada. Vet Res 48, 10. doi:10.1186/s13567-017-0417-6

Zheng, X., Zheng, H., Lan, R., Ye, C., Wang, Y., Zhang, J., Jing, H., Chen, C., Segura, M., Gottschalk, M., Xu, J., 2011. Identification of genes and genomic islands correlated with high pathogenicity in *Streptococcus suis* using whole genome tilling microarrays. PLoS ONE 6, e17987. doi:10.1371/journal.pone.0017987

Zhu, W., Wu, C., Sun, X., Zhang, A., Zhu, J., Hua, Y., Chen, H., Jin, M., 2013. Characterization of *Streptococcus suis* serotype 2 isolates from China. Veterinary Microbiology 166, 527–534. doi:10.1016/j.vetmic.2013.06.009

Zou, G., Zhou, J., Xiao, R., Zhang, Liangsheng, Cheng, Y., Jin, H., Li, L., Zhang, Lijun, Wu, B., Qian, P., Li, S., Ren, L., Wang, J., Oshota, O., Hernandez-Garcia, J., Wileman, T.M., Bentley, S., Weinert, L., Maskell, D.J., Tucker, A.W. (Dan), Zhou, R., 2018. Effects of environmental and management-associated factors on prevalence and diversity of *Streptococcus suis* in clinically healthy pig herds in China and the United Kingdom. Appl Environ Microbiol 84, e02590-17. doi:10.1128/AEM.02590-17

APPENDIX

Appendix 4.1: Genome assembly and annotation statistics of S. suis isolates (n=208).

UniqueID	SRR accession	Total length (bp)	No. contigs	GC (%)	Avg. genome coverage	No. protein coding sequences	N50	L50
2015UMN3713.8	SRR9123061	2133653	87	41.13	172	2064	47393	15
2015UMN3791.4	SRR9123062	2169901	61	41.02	171	2093	76636	8
2014UMN2806.2	SRR9123063	2224364	90	40.95	163	2151	51645	12
2014UMN3951.1	SRR9123064	2241982	85	41.17	172	2100	59849	11
2014UMN4255.1	SRR9123065	2416440	223	41.32	177	2399	22465	32
2015UMN43.3	SRR9123066	2092370	60	41.15	196	2000	80070	10
2016UMN917.8	SRR9123067	2101247	56	41.14	214	2005	80351	10
2015UMN3425.7	SRR9123068	2170891	56	41.01	179	2094	81926	8
2016UMN1470.8	SRR9123069	2184186	62	41.07	203	2151	69317	11
2016UMN1440.1	SRR9123070	2199101	145	40.97	211	2157	26227	29
2015UMN3851.7	SRR9123071	2028878	42	41.17	224	1928	92659	7
2015UMN4009.2	SRR9123072	2019660	96	41.25	209	1914	44021	16
2015UMN3832.5	SRR9123073	2170615	56	41.01	198	2098	81900	8
2015UMN3849.1	SRR9123074	2137131	72	41.1	220	2073	63045	12
2016UMN102.5	SRR9123075	2096333	59	41.15	208	2004	77848	10
2016UMN125.4	SRR9123076	2028881	47	41.42	223	1865	122409	6
2017UMN1562.91	SRR9123077	2211674	88	41.18	230	2050	75915	10
2017KSU2054.24	SRR9123078	2112418	110	41.22	239	2025	39848	14
2017UMN1574.421	SRR9123079	2040766	102	41.32	266	1955	43796	14
2017UMN1562.912	SRR9123080	2407512	146	41.29	191	2346	31473	23
2017UMN1464.2	SRR9123081	2139699	93	41.16	266	2113	38178	17
2017UMN1464.4	SRR9123082	2147652	87	41.14	296	2125	40175	16
2017UMN1562.93	SRR9123084	2002592	112	41.55	222	1905	38729	18
2017UMN1562.911	SRR9123085	2149228	62	41.16	247	1979	113271	6
2017UMN1562.95	SRR9123086	2405645	166	41.43	165	2346	32764	25
2014UMN3704.91B	SRR9123087	2102433	57	41.16	198	2006	75592	10
2014UMN3704.91A	SRR9123088	2099068	65	41.18	187	2002	57370	11
2014UMN3631.1	SRR9123089	2203530	52	41.07	217	2145	91778	9
2014UMN3368.9	SRR9123090	2082904	71	41.15	221	1981	61658	12
2015UMN63	SRR9123091	2024303	82	41.52	202	1985	73390	10

2014UMN4309.1	SRR9123092	2138539	71	41.12	189	2105	63042	13
2014UMN3038.4	SRR9123093	2305564	55	40.94	194	2311	103353	9
2014UMN1785.4	SRR9123094	2175372	72	41.05	168	2154	66862	12
2014UMN3602.9	SRR9123095	2063805	106	41.18	214	2030	33234	18
2014UMN3212	SRR9123096	2122961	55	41.09	208	2069	67579	10
2015UMN56.5	SRR9123097	2136532	71	41.06	220	2083	61304	13
2015UMN1614.1	SRR9123098	2163030	84	41.07	170	2107	66109	12
2016UMN209.8	SRR9123099	2144840	81	41.09	223	2085	55302	15
2016UMN146.4	SRR9123100	2119510	75	41.12	227	2075	72976	11
2016UMN362.1	SRR9123101	2184262	71	41.03	201	2149	68198	10
2016UMN274.7	SRR9123102	2134295	78	41.14	206	2104	43906	16
2016UMN578.4	SRR9123103	2241933	69	40.89	260	2181	96464	9
2016UMN462.4	SRR9123104	1963774	83	41.44	246	1935	51235	13
2016UMN835.2	SRR9123105	2127607	65	40.93	258	2023	91279	9
2016UMN668.5	SRR9123106	2023234	31	41.22	194	1926	129396	5
2016UMN924.3	SRR9123107	2234671	84	40.95	206	2157	64035	11
2016UMN884.2	SRR9123108	2133522	78	41.14	216	2057	52237	12
2014KSU2091.1	SRR9123109	2161751	61	41.34	287	2023	67967	10
2014KSU939.9	SRR9123110	2017386	36	41.25	265	1926	118548	5
2017KSU2087.212T	SRR9123111	2391036	156	41.3	183	2318	29307	25
2017KSU2054.29R	SRR9123112	2373909	145	41.24	223	2259	33917	24
2017KSU2087.221AT	SRR9123113	2198471	44	41.11	241	2205	138187	6
2017KSU2087.212V	SRR9123114	2393612	154	41.3	216	2319	32397	22
2017KSU1243.9	SRR9123115	2114006	56	41.13	247	2059	80835	9
2017KSU220.2	SRR9123116	2073537	49	41.34	268	1976	88722	8
2014KSU27.8	SRR9123117	2128200	75	41.32	220	2029	56403	11
2017KSU1528.8	SRR9123118	2136328	86	41.19	255	2089	69605	9
2014UMN3794.6	SRR9123119	2109567	76	41.14	168	2014	64050	12
2014UMN3978.4	SRR9123120	2120359	66	41.09	226	2073	66661	11
2015UMN647.4	SRR9123121	2245951	89	40.99	166	2229	50494	10
2015UMN1735.7	SRR9123122	2401604	90	41.05	192	2321	54798	16
2015UMN344.2	SRR9123123	2450758	174	41.23	109	2395	25954	32
2015UMN638.1	SRR9123124	2105077	64	41.14	177	2054	65873	12
2015UMN64.3	SRR9123125	2411509	158	41.17	160	2335	26490	27
2015UMN277.7	SRR9123126	2110326	59	41.06	210	2009	100845	6
2015UMN1499.5	SRR9123127	2141587	77	41.15	181	2117	51723	14
2014UMN4352.4	SRR9123128	1968827	128	41.48	170	1942	30124	21
2014UMN3634.9	SRR9123129	2141343	79	41.15	163	2115	62432	14

2014UMN3624.71	SRR9123130	2084029	58	41.26	203	2022	72808	11
2014UMN1700.2	SRR9123131	2226070	84	41.02	164	2216	63051	13
2014UMN1245	SRR9123132	2183940	102	41.18	164	2124	58697	12
2016UMN2885.2	SRR9123133	2161901	70	41.04	239	2128	72885	10
2016UMN2940.9	SRR9123134	2097059	64	41.15	171	2003	82390	10
2016UMN2677.3	SRR9123135	2112850	52	41.1	233	2070	80590	9
2016UMN2813.1	SRR9123136	2042805	58	41.25	270	1953	77041	9
2016UMN2628.1	SRR9123137	2052010	57	41.26	139	1987	95544	8
2016UMN2642.9	SRR9123138	2166581	66	41.03	243	2134	75341	10
2017UMN631.2	SRR9123139	2040444	82	41.26	216	1949	52970	13
2016UMN2627.9	SRR9123140	2052611	57	41.26	257	1986	76697	9
2017UMN540.4	SRR9123141	2092379	68	41.16	257	1994	61345	13
2017UMN564.5	SRR9123142	2140566	78	41.17	292	2110	60562	14
2016UMN1917.1	SRR9123143	2043833	96	41.36	210	2014	45090	13
2016UMN1250.1	SRR9123144	2113187	117	40.98	188	1999	52820	12
2015UMN4053.5	SRR9123145	2161052	56	41.03	198	2102	85422	9
2015UMN4083.1	SRR9123146	2150305	63	41.12	208	2128	67567	12
2015UMN4145.9	SRR9123147	2088616	59	41.16	226	1992	80351	10
2016UMN574.9	SRR9123148	2217752	70	40.88	203	2162	75366	11
2015UMN3479.3	SRR9123149	2056255	40	41.23	212	1994	95594	6
2015UMN3767.2	SRR9123150	2134669	65	41.13	215	2107	67617	11
2015UMN3812.4	SRR9123151	2120916	39	41.01	170	2077	77385	7
2015UMN3886.2	SRR9123152	2106236	60	41.14	184	2057	80555	9
2016UMN3248.4	SRR9123153	2070581	77	41.21	249	1975	53864	14
2017UMN1574.428	SRR9123154	2412133	125	41.3	221	2346	41599	18
2016UMN3117.3	SRR9123155	2119061	26	41.19	260	1984	260919	3
2016UMN3106.8	SRR9123156	2244112	69	40.97	245	2221	61894	11
2016UMN3242.3	SRR9123157	2169958	62	41.01	187	2096	80192	9
2016UMN3120.1	SRR9123158	2060835	39	41.19	256	1995	119634	6
2016UMN2965.6	SRR9123159	2052180	45	41.25	277	1986	103976	7
2016UMN2948.1	SRR9123160	2155401	70	41.05	232	2126	71723	11
2016UMN3047	SRR9123161	2079770	57	41.19	236	2032	66773	12
2016UMN3027.6	SRR9123162	2109665	68	41.15	288	2064	67542	12
2014UMN3720.5	SRR9123163	2106351	59	41.1	195	2056	66628	11
2014UMN3930.3	SRR9123164	2125476	64	41.15	65	2061	101438	8
2015UMN147.1	SRR9123165	2129931	83	41.17	180	2053	44719	15
2015UMN865.5	SRR9123166	2141926	59	41.11	184	2113	73707	11
2015UMN682.6	SRR9123167	2161819	97	41.08	230	2103	41471	14

2014UMN1803	SRR9123168	2137122	116	41.2	163	2105	29774	21
2014UMN2149	SRR9123169	2101431	39	41.26	180	2054	119532	7
2014UMN3295.5	SRR9123170	2072110	75	41.2	203	2029	63137	12
2016UMN763.3	SRR9123171	2228371	113	40.89	175	2153	76292	10
2016UMN858.4	SRR9123172	2056642	39	41.23	150	1997	103996	6
2016UMN2500.62B	SRR9123173	2074220	123	41.17	236	1992	33187	23
2017UMN355.91	SRR9123174	2110762	81	41.14	283	1997	57687	13
2016UMN1245.7	SRR9123175	2150269	73	41.12	206	2125	63189	13
2016UMN1367.1	SRR9123176	2047385	66	41.26	220	1961	74102	11
2016UMN1452.1	SRR9123177	2071333	34	41.37	187	1924	142166	6
2015UMN3869.3	SRR9123178	2239872	80	40.93	230	2165	76261	10
2016UMN2465.63A	SRR9123179	2301650	49	40.84	191	2308	123888	8
2016UMN2465.63B	SRR9123180	2053443	56	41.07	240	2010	63088	10
2016UMN2500.63	SRR9123181	2076066	43	41.05	192	1978	129611	6
2016UMN2500.62A	SRR9123182	2073378	120	41.17	217	1990	33858	23
2015UMN2869.3	SRR9123183	2171471	57	41.02	179	2096	66771	9
2015UMN2795.3	SRR9123184	2124792	70	41.03	170	2009	64658	10
2015UMN1613.2	SRR9123185	2152201	72	41.14	141	2131	61852	13
2015UMN1502.7	SRR9123186	2120054	73	41.12	149	2082	63251	12
2015UMN1933.2	SRR9123187	2141803	71	41.15	143	2114	50682	14
2015UMN1635.9	SRR9123188	2163934	59	41.22	158	2142	75521	10
2015UMN2430.5	SRR9123189	1952168	55	41.35	156	1854	73970	9
2015UMN2400.6	SRR9123190	1972371	31	41.4	197	1897	120217	5
2015UMN2626.7	SRR9123191	2183732	55	41.01	163	2115	74913	11
2015UMN2473.3	SRR9123192	2079133	91	41.16	163	1980	50741	14
2014UMN3749.9	SRR9123193	2109815	75	41.12	145	2014	75384	10
2014UMN3627	SRR9123194	2142769	73	41.14	181	2113	63120	13
2014UMN3243.9	SRR9123195	2145257	125	41.34	104	2071	30843	20
2014UMN2470.9	SRR9123196	2312752	123	41.44	177	2186	38099	21
2015UMN1919.5	SRR9123197	2229125	99	40.96	175	2169	66126	10
2015UMN745.2	SRR9123198	2328621	135	41.51	147	2215	38112	22
2015UMN370.7	SRR9123199	2164307	79	41.07	208	2108	66628	11
2015UMN370.6	SRR9123200	2397694	94	41.56	170	2272	54304	14
2015UMN367.3	SRR9123201	2138103	74	41.11	215	2083	55200	13
2015UMN302.5	SRR9123202	2111029	66	41.16	208	2084	71722	10
2017UMN236.4	SRR9123203	2260763	67	41.01	228	2260	104107	7
2017UMN74.9	SRR9123204	2140173	77	41.17	249	2112	60562	14
2016UMN1524.94	SRR9123205	2349829	157	41.47	246	2248	27788	25

2016UMN1524.93	SRR9123206	2005990	53	41.37	318	1943	71696	8
2017UMN1435.22	SRR9123207	2107190	29	41.47	277	1971	205235	4
2017UMN1435.21	SRR9123208	2146410	102	41.2	224	2077	49940	15
2017UMN937.8	SRR9123209	2149814	73	41.14	302	2003	58327	11
2017UMN355.92	SRR9123210	2076642	15	41.3	272	1946	327963	3
2016KSU16.8	SRR9123211	2154414	81	41.13	283	2106	57371	13
2015KSU2327.2	SRR9123212	2163061	57	41.35	270	2030	74081	9
2016KSU697.6	SRR9123213	2095507	80	41.17	197	2045	54734	12
2016KSU47.5	SRR9123214	2208929	51	40.91	261	2172	152118	6
2017UMN1562.92B	SRR9123215	2398283	178	41.45	161	2336	28907	27
2017UMN524.8	SRR9123217	2117117	65	41.14	248	2076	62416	11
2017UMN416.5	SRR9123218	2137640	86	41.17	266	2110	51032	13
2015UMN1018.1	SRR9123219	2210117	71	41.04	164	2171	55281	12
2015UMN1502.6	SRR9123220	2122679	62	41.09	143	2088	67679	13
2015UMN592.5	SRR9123221	2146547	59	41.1	197	2118	80407	9
2015UMN637.3	SRR9123222	2152044	79	41.12	199	2106	50650	14
2015UMN122.1	SRR9123223	2088904	62	41.14	213	1996	77852	11
2015UMN138	SRR9123224	2084808	67	41.16	174	1983	61986	12
2015UMN856	SRR9123225	2141782	66	41.14	199	2114	63189	13
2015UMN890.4	SRR9123226	2165776	73	41.01	158	2106	63042	11
2015UMN689	SRR9123227	2073216	82	41.2	167	2029	47033	14
2015UMN720.1	SRR9123228	2181035	48	41.03	188	2118	100845	7
2015UMN193.2	SRR9123229	2177436	111	40.98	177	2092	50374	12
2015UMN193.3	SRR9123230	2170533	41	41.02	185	2126	127296	6
2015UMN3222.1	SRR9123231	2196422	82	40.94	197	2097	64580	11
2015UMN3436.5	SRR9123232	2174324	84	41.06	159	2149	51682	14
2014UMN2228.9	SRR9123233	2041137	61	41.29	170	1950	70847	10
2014UMN2148.7	SRR9123234	2112649	95	41.04	188	2064	32748	20
2014UMN2784.1	SRR9123235	1963631	78	41.44	68	1937	51236	12
2014UMN2436.2	SRR9123236	2150682	71	41.12	163	2127	63189	13
2014UMN3624.74	SRR9123237	2121635	57	41.05	225	2079	70605	9
2014UMN3342.6	SRR9123238	2120569	79	41.22	204	2099	55166	14
2017UMN92.8	SRR9123239	2304127	81	40.76	269	2246	60139	16
2017UMN144.4	SRR9123240	2039646	87	41.3	214	1948	58785	12
2017UMN145.1	SRR9123241	2110421	58	41.11	324	2011	97569	8
2017UMN164.7	SRR9123242	2109280	53	41.29	244	1975	84844	9
2016UMN3250.3	SRR9123243	2091935	61	41.14	220	1998	74643	9
2017UMN52.2	SRR9123244	2047603	31	41.07	220	1943	125460	5

2017UMN55.9	SRR9123245	2084953	55	41.26	269	2022	87201	9
2017UMN92.6	SRR9123246	2135788	66	41.13	266	2104	63201	12
2017UMN196	SRR9123247	2104094	57	41.12	268	2051	72225	10
2017UMN266.4	SRR9123248	2290451	52	40.9	226	2290	84606	7
2014UMN3928.6	SRR9123249	2089700	56	41.13	197	1994	80466	9
2014UMN3787.2	SRR9123250	2095657	80	41.18	191	2039	47343	15
2014UMN3786.4	SRR9123251	2215337	128	41.13	202	2153	33175	18
2014UMN3714.6	SRR9123252	2133238	65	41.04	170	2089	65685	11
2015UMN27.5	SRR9123253	2119784	84	41.16	194	2039	50775	15
2014UMN4118.6	SRR9123254	2060292	80	41.21	236	2016	47541	12
2014UMN4033.5	SRR9123255	2153974	67	41.14	215	2134	62611	12
2014UMN4007.5	SRR9123256	2104869	66	41.14	183	2054	75309	10
2015UMN80.9	SRR9123257	2029198	34	41.17	207	1927	129563	6
2015UMN36	SRR9123258	2051596	56	41.27	205	1988	81784	8
2017UMN583.4	SRR9123259	2207148	71	41.18	255	2140	81650	10
2017UMN539.9	SRR9123260	2166887	112	41.11	215	2130	48052	14
2017UMN500.8	SRR9123261	2082771	78	41.16	239	1978	55190	14
2017UMN296.1	SRR9123262	2114516	71	41.05	287	2014	63719	12
2017UMN1049.2	SRR9123263	2045217	80	41.23	235	1953	61006	12
2017UMN1049.1	SRR9123264	2179209	111	40.99	253	2073	58398	12
2017UMN991.6	SRR9123265	2163161	60	41	182	2093	85422	8
2017UMN834.1	SRR9123266	2108145	73	41.19	261	2079	53081	12
2017UMN1417.8	SRR9123267	2108632	118	41.18	207	2048	36369	19
2017UMN1161.8	SRR9123268	2241865	142	41.08	240	2198	41104	17